

## Gathering the evidence on medical marijuana

**The US Supreme Court has been hearing arguments about whether states that have legalized the medical use of marijuana are in breach of federal law. But medical agencies engaged in its clinical testing should not be deterred from their work.**

It has been a long war of attrition, but patient groups, who sometimes know a thing or two about treatment efficacy, have started to persuade governments and medical research agencies to confront the issue of whether marijuana has medical value.

Thousands of years of marijuana use suggest that it may have. But since the 1960s, marijuana has been classified by international agencies as a 'schedule 1' drug; that is, an addictive substance with no recognized medical value. This illegal status has, perversely, hindered scientists from conducting the research that might determine whether either of these attributes is true.

Many patients say that smoking the weed makes them feel better. Among them are cancer and AIDS sufferers, who claim that it helps them control the nausea and vomiting that accompany chemotherapy, relieves pain resistant even to morphine and, as an important bonus, restores appetite. Multiple sclerosis sufferers claim relief from muscle spasm. A little accompanying mood elevation is not necessarily an evil, in the circumstances, they say — and some of their physicians agree with them.

There is also a biological basis for these therapeutic claims. Research has shown that the active constituents of marijuana — the cannabinoids — operate through cannabinoid receptors present throughout the nervous system. These are normally activated by endogenous cannabinoids, much as endorphins operate through opiate receptors, through which morphine and heroin exert their effects.

Nonetheless, governments, including the US federal government, have until recently refused to sanction the medical use of marijuana, and have also done what they can to prevent its clinical testing. They have defended their inaction by claiming that either step would signal to the public a softening of the so-called 'war on drugs'.

But the debate is slowly edging forward. In a recent referendum,

Californians voted to legalize the medical use of marijuana, and several other states have followed suit. Medical research agencies in the United Kingdom and Canada are sponsoring clinical trials (see *Nature* **410**, 505; 2001). Such trials will help to define the medical disorders that marijuana might usefully treat, and to determine the therapeutic value of its most potent ingredient, delta-9 tetrahydrocannabinol (THC).

The illegal status of marijuana has endowed it with mystical properties. Many regard marijuana as a sort of panacea, withheld from the public by politics and by an indifferent pharmaceutical industry which sees no profit in developing an unpatentable weed into medicine. Some view synthetic THC with suspicion, believing that the natural mixture of cannabinoids in the marijuana leaf appropriately balances therapeutic effects and side effects.

This approach is unhelpful. The drug may have a wide range of therapeutic uses, but for some of these, including nausea and vomiting and glaucoma, better alternatives are already available. For others, such as spastic cramps and the control of neurogenic pain, marijuana-based drugs might prove to be genuinely unique and important.

These uses should be investigated by classical pharmacological methods, so that efficacy can be established — and optimized — unequivocally. That means working with pure compounds. After all, new drugs derived from morphine are clearly better for treating severe pain than are extracts of opium poppies. If the ingredients of marijuana prove to be effective, smoking of the natural substance will be replaced by a more refined product.

The pharmacology of cannabinoids is a valid field of scientific investigation. Pharmacologists have the tools and the methodologies to realize its considerable potential, provided the political climate permits them to do so. ■

## E-access to science

**Introducing a forum on the future of primary scientific publishing.**

This week, *Nature* launches an online forum on a topic that has filled volumes of conference proceedings and reams of individual articles since the emergence of the Internet — namely, the web's impact on the publishing of the results of original research.

The most recent and prominent manifestation of the debates surrounding this topic is an initiative by researchers to force publishers, by a threatened boycott, to release archived reports of original research into centralized, freely available and unrestricted databases, known as 'The Public Library of Science' (PLS). *Nature's* forum does not represent the response of our publishers, the Nature Publishing Group, to the PLS initiative. *Nature's* publishers are currently soliciting feedback from researchers, librarians and other interested parties in weighing up the issues. But, in principle, everyone in research has

an interest in understanding the many aspects of those issues. So the focus of the forum is only partly on the debate between, for example, advocates of free access and those who worry about the loss of publishers' livelihood and ownership rights.

Accordingly, we have commissioned not only researchers and publishers but also experts in scientific information management and commerce to contribute. This week you can find the first contributions, including two views from the European Molecular Biology Organization, which is proposing its own model, to be funded by the European Union (see <http://www.nature.com/nature/debates/e-access>). We welcome original responses to the forum, bearing in mind our interest in exploring in more depth the many aspects of the online-access debate, rather than reiterating opinions on how that access should be shaped. ■





**Bush under fire**  
Disappointment  
greeted US exit from  
Kyoto Protocol  
p616



**Double trouble**  
Congress considers  
ban on human  
cloning  
p617



**Green block**  
Ecologists stall land  
reclamation project  
in Japan  
p619



**Grounded**  
India's space plans  
in disarray as rocket  
fails to fly  
p619

# Experts clash over likely impact of cheap AIDS drugs in Africa

**Meredith Wadman, Washington**

Conflict is brewing between clinicians treating AIDS patients and some public-health experts over the wisdom of distributing cut-price AIDS drugs rapidly in poor countries.

While hard-pressed doctors are trying to get the medicines out as quickly as possible, others warn that they may do more harm than good, diverting scarce funds from more cost-effective preventive measures and perhaps raising new problems with drug resistance and side effects.

"We have to ask what capabilities need to be put in place to make sure the problem isn't made worse," says Tom Coates, director of the AIDS Research Institute at the University of California, San Francisco.

"Parachuting drugs into countries and saying, 'Here they are, go for it!' could do a considerable amount of harm in addition to the good we're expecting," says Tony Fauci, director of the National Institute of Allergy and Infectious Diseases.

The warnings follow the decision last



Hard to swallow: combination pills that make AIDS drug regimens easier to follow are on the way.

month by several leading drug companies to slash the prices of AIDS drugs for sub-Saharan Africa, home to 70% of the world's HIV-infected people (see *Nature* 410, 289; 2001). The new prices followed moves by generic-drug makers in India and elsewhere to defy

patents and distribute AIDS drug cocktails in Africa (see *Nature* 409, 751; 2001).

These actions created an unprecedented momentum for the wide distribution of AIDS medicines on a continent where, just a few months ago, no one dreamt of broad access to expensive, life-saving therapies.

But some public-health experts worry that the arrival of more of the drugs will eat into crucial AIDS-prevention budgets. They also say that the infrastructure in many places will not support proper drug distribution, and that bad distribution and monitoring may give rise to strains of the virus that will resist known therapies.

Barry Bloom, dean of the Harvard School of Public Health, says AIDS drugs should be dispensed only under direct observation — a strategy that has worked for tuberculosis — and that may not only prevent resistance, but also stop black markets from developing.

These arguments inflame doctors and activists working with AIDS patients in poor countries. The warnings are reminiscent, they say, of the now-debunked criticism of widescale immunization initiatives a generation ago. Anne-Valerie Kaninda, a medical adviser to Doctors Without Borders (Médecins sans Frontières) in New York, says: "The limited experience of introducing anti-retrovirals [in poor countries] shows that there are solutions to the problems."

One case study in particular backs up

## Farmer fined for growing GM seed

**David Spurgeon, Montreal**

A Canadian judge has ruled that a farmer must pay royalties to the biotechnology giant Monsanto for growing its genetically modified seed without the supplier's permission.

Percy Schmeiser, a 70-year-old grain farmer from Bruno, Saskatchewan, says the Roundup Ready canola (oilseed rape) seed found growing on his land was the result of contamination by pollen from neighbouring fields or seed blown off passing trucks. Monsanto said he knowingly planted it without paying for it.

But without ruling directly on that question, Justice Andrew MacKay said Schmeiser had infringed Monsanto's licences, and gave both sets of lawyers three weeks to settle appropriate damages. If they fail, he has ruled, Schmeiser must pay Monsanto a licence fee of Can\$15,450 (US\$9,800) — not the Can\$105,000

demand by the company. Moreover, the judge declined to order additional exemplary damages.

Schmeiser's lawyer, Terry Zakreski, argued that Monsanto had lost its exclusive rights to its patent on the gene — which allows plants to survive the application of Monsanto's weedkiller Roundup — when it was released into the environment. He said farmers are allowed to keep seed from their crops to plant the following year, and although Monsanto requires farmers using its canola seeds to sign an agreement that they will not do so, the company has no such power over farmers who do not buy Roundup Ready canola.

Trish Jordan, a spokeswoman for Monsanto in Canada, said the decision will help to protect the property rights of the company and of the thousands of farmers who pay for its technology.



their argument. Four years ago the Brazilian government began providing free cocktails of three AIDS drugs. Despite an imperfect public-health system, the country's AIDS death rate has been halved. Last year, the number of new infections was half of that predicted by the World Bank in 1994.

"It's horribly difficult to make [drug distribution] happen. But it is possible," says Ezio Santos Filho, a Brazilian activist whose group, Grupo Pela Vidda, has helped push for the programme. "What we cannot accept is the excuse that the regimens are just too difficult" for Africans. "We have to make these medications available by all means possible."

The drug companies are playing down the effect their move will have without major infrastructure improvements. "There's a need for more AIDS specialists to make sure these patients take their medicines properly every day, so that we avoid resistance developing and the medicines are truly effective," says Jeff Trewitt, a spokesman for the Pharmaceutical Research and Manufacturers of America.

But advocates of access to the drugs point out that, even in the United States, adherence to complicated drug regimens is hardly a given: recent studies have shown that only two-thirds of people with AIDS take their medicine as much as 80% of the time. The number in Brazil is about the same, according to a 1999 government study, providing ammunition for those who say the real enemy of getting medicines to Africa is a lack of political will, rather than practical obstacles.

Proponents of widespread distribution also argue that the complicated drug regimens used in rich nations may not be needed in poorer countries. A pilot programme being launched in South Africa this month by Doctors Without Borders will have patients taking three pills twice a day, rather than the 20 or 30 tablets commonly taken in developed countries.

Follow-up tests will be administered less frequently than in the developed world, unless patients show visible deterioration. Ultimately, simpler and cheaper versions of the tests are expected to be produced. And combination pills that make drug regimens easier to follow are already being produced by some drug companies.

Eric Goemaere, the Doctors Without Borders doctor in charge of the pilot programme in Khaelitsha township near Cape Town, admits that this approach to monitoring has not been scientifically validated. "What we need to develop is a science for low-income countries," he says. But he adds that he is not prepared to wait for that science to emerge before starting treatment. "We see [people] die almost on a daily basis. We feel a bit in a hurry." ■

## Criticism mounts as Bush backs out of Kyoto accord

**Mark Schrope**

Leading US climate scientists are disappointed over President George W. Bush's decision to abandon the Kyoto Protocol for reducing greenhouse-gas emissions.

But unlike Bush's political opponents, who vehemently condemned the move, the researchers still hope Bush will act on climate change once he has appointed key officials.

"I think it would be very unfortunate to scrap the Kyoto Protocol at this point," says James McCarthy, a Harvard oceanographer and co-chair of the Intergovernmental Panel on Climate Change's working group on impacts, adaptation and vulnerability.

McCarthy says a key problem is that Bush has yet to fill all of the main scientific advisory positions in his administration. Sherwood Rowland, an atmospheric scientist at the University of California, Irvine, who shared the 1995 chemistry Nobel prize for his work on ozone, agrees. "The absence of any high-level scientific advice is quite noticeable," he says.

Henry Jacoby, an environmental economist at the Massachusetts Institute of Technology, believes the Bush administration will find a coherent approach to climate change. "Kyoto is a set of numbers and conditions they don't like, but it is also a process, and it is not clear they're rejecting the process," he says.

Before meeting with German Chancellor Gerhard Schröder in Washington last week, Bush said that he would not accept any plan that harms the US economy, as he believes the Kyoto treaty does. After the meeting, Schröder said he disagreed with Bush, but that the United States must make its own decision on climate-change policy.



Agreeing to disagree: Bush (right) and Schröder share a joke on their way to last week's meeting.

White House spokesman Ari Fleischer said repeatedly last week that the administration felt no obligation to the treaty because the Senate had rejected it. But the Senate has never voted on the Kyoto Protocol itself.

Senator Joseph Lieberman (Democrat, Connecticut) announced last week that he would investigate the methods used by the Bush administration to make environmental decisions, saying that they "ignore the public interest and defy common sense".

International representatives will meet in Bonn in July to discuss the Kyoto accord. ■

## Climate change transforms island ecosystem

**Peter Pockley, Sydney**

The first survey for a decade of animals and plants on Australia's Heard Island, 4,000 kilometres southwest of Perth, has unearthed dramatic evidence of global warming's ecological impact.

The survey found that glaciers



Getting warmer: botanists hunt for clues on Heard Island.

on the sub-Antarctic island have retreated by 12% since the first measurements were taken in 1947. This has been accompanied by a rise in sea surface temperature of up to 1 °C and rapid increases in flora and fauna.

Dana Bergstrom, an ecologist at the University of Queensland in Brisbane, who led the survey's study of plant life, says areas that were previously poorly vegetated are now "lush with large expanses of plants".

Populations of birds, fur seals and insects have also expanded rapidly since earlier studies, says Eric Woehler of the Australian Antarctic Division at Kingston,

Tasmania, part of Australia's environment department. According to Woehler, the number of king penguins has exploded from only three breeding pairs in 1947 to 25,000, while the Heard Island cormorant, listed previously as "vulnerable", has increased to 1,200 pairs. From near extinction, fur seals now number 28,000 adults and 1,000 pups.

Australia has now committed to surveying Heard Island every two years for the next decade as part of an international programme to measure the response of ecosystems to climate change in the Antarctic. ■



# Fears of cults and kooks push Congress towards cloning ban

Paul Smaglik, Washington

The United States is moving closer to outlawing human reproductive cloning after a dramatic congressional hearing drew attention to the prospect that attempts to clone a human are imminent.

But despite agreement on Capitol Hill that such experiments should be banned, the means of achieving a ban remain elusive. Scientific societies fear that any proposed legislation might be amended by opponents of abortion to ban not only human cloning, but also therapeutic cloning and other scientifically useful work.

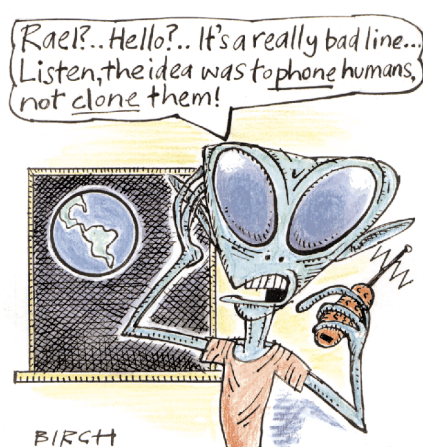
Billy Tauzin (Republican, Louisiana), chair of the House Committee on Energy and Commerce, whose investigations subcommittee held the hearing on 28 March, said before the hearing that he had not decided whether to propose legislation. But several committee members said during the hearing that legislation was necessary, and Tauzin's staff said later that he may introduce a bill when Congress reconvenes after Easter.

The hearing was triggered by claims from two research groups that each intends to clone a human soon. Both groups were asked to testify at the hearing, despite some scientists' concerns that this would alarm and confuse the public.

A company called Clonaid, founded by a man who claims extraterrestrials told him to change his name to Raël, and whose parent group, Valiant Venture Ltd, is registered in the Bahamas, says it plans to clone a human at an undisclosed location in the United States. Raël told the hearing that the human race had been cloned by extraterrestrials and that it was our duty to clone ourselves.

An international group of reproduction specialists, meanwhile, has announced plans to use human cloning as a technique to help infertile couples have children (see *Nature* 410, 293; 2001). Neither Brigitte Boisselier, Clonaid's scientific director, nor Panos Michael Zavos, a partner in the other effort, has ever successfully cloned an animal.

Other witnesses sought to draw distinctions between good and bad science and between safety and ethics. Congress's perception of such distinctions is likely to shape any legislation or regulation of cloning. A focus on safety, rather than ethics, might provide a basis for regulation by the Food and Drug Administration, the hearing was told.



Arthur Caplan, head of the Center for Bioethics at the University of Pennsylvania, told the hearing that immediate safety concerns should supercede any ethical discussion. Such concerns are so great, he said, that only "cults, cranks, kooks and capitalists" would propose cloning a human given the known risks.

Thomas Okarma, president of the California-based biopharmaceutical company Geron, testified on behalf of the Biotechnology Industry Organization. He said human reproductive cloning should be banned for safety reasons, because the success rate in animals is between 3 and 5%.

Boisselier claimed that success rates in animals were no indication of the technique's efficacy in humans. And Zavos said his group would pre-screen the viability of embryos before implanting them.

Rudolf Jaenisch, a professor of biology at the Massachusetts Institute of Technology, said these arguments contained "serious factual errors", the most glaring of which was the claim that the researchers could screen for viable embryos. This would mean looking at the expression of all 30,000 or so human genes, which is currently not technically possible in any organism, Jaenisch said.

Jaenisch added that even if such technology existed, using it on an embryo would not help, because many genes are expressed only later in development. After the hearing, he said he was concerned that airing the claims of the two groups could taint mainstream science. "These nuts get much too much publicity," he said.

Larry Goldstein of the University of California at San Diego, vice-chairman of the American Society for Cell Biology public policy committee, says he is concerned that such publicity might lead to hasty legislation that would block potentially valuable areas of research.

## Bush appoints venture capitalist as technology adviser

Irwin Goodwin, Washington

President George W. Bush has surprised scientists with the appointment of his first main adviser on science and technology.

Bush last week named Floyd Kvamme as co-chairman of the President's Committee of Advisors on Science and Technology (PCAST).

Kvamme is an electrical engineer with a management track record in several Silicon Valley corporations, including National Semiconductor and Apple, and is a partner in the venture-capital firm of Kleiner Perkins Caufield & Byers in Menlo Park, California. But he is best known in Washington as chairman of the conservative advocacy group Empower America.

Announcing the appointment of Kvamme, Bush said: "He is an entrepreneur. He is a risk taker. He understands risk and reward. But, more importantly, he knows the players, the people who can bring good, sound advice to this administration."

Bush's decision to appoint Kvamme before filling the post of president's science adviser — who leads the White House Office of Science and Technology Policy (OSTP) and will be PCAST's other co-chair — is seen as unusual by some. In addition, the White House did not consult scientific or engineering societies, or the national academies, which like to provide lists of suitable candidates.

"It's most unusual to name a co-chair of PCAST without having the key figures of the OSTP on board," says Jack Gibbons, who was science adviser to former president Bill Clinton.



Techno speak: Bush anticipates 'good, sound advice' from appointee Floyd Kvamme.

# Madrid quakes as Barcelona shows ambition

**Xavier Bosch, Barcelona**

Madrid's traditional dominance of Spanish science is coming under pressure from the growing aspirations of Barcelona, its great historical rival. And scientists in the Spanish capital are concerned that their Catalanian competitor will beat them in terms of funding for research initiatives.

Madrid's researchers note that, since last spring, two Catalan ministers have been in charge of the Ministry of Science and Technology (MST), which controls the flow of research funds from the Spanish government. The appointments were seen as an attempt by the conservative government to shore up its weak support in Catalonia.

Researchers in Barcelona have long complained about Madrid's dominance of Spanish science. They point out, for example, that during Franco's rule the Higher Council of Scientific Research (CSIC) — the country's largest research agency — built most of its research centres in Madrid.

Despite moves to decentralize Spain after the dictator's death in 1975 — including the establishment of a powerful regional government for Catalonia — research activity remains heavily concentrated in the capital. In 1999, total research spending in



**Andreu Mas-Colell**  
wants to distribute  
research around Spain.

Madrid was Pta 265 billion (US\$1.4 billion), around 1.5% of the economy, against Pta 188 billion (1.1% of the economy) in Catalonia.

"There's a serious research deficit in Barcelona," says Pere Puigdomènech, who is director of the city's

Institute of Molecular Biology. He says two-thirds of CSIC biomedical researchers are located in Madrid. "The government has transferred practically no research powers to Catalonia," he says.

Nevertheless, the Catalan regional government set up a research department last April with an annual budget of Pta 8.5 billion. And Barcelona scientists have established several successful research centres, including the Municipal Institute of Medical Research and the Institute for High Energy Physics.

But some recent developments have put Madrid's scientists on the defensive. For example, the MST has offered Barcelona Pta 5 billion to build a science park, against Pta 2.8 billion for a similar project in Madrid.

Rodolfo Miranda, vice-chancellor of the Autonomous University of Madrid, accepts this distribution in funds. But he is annoyed that Barcelona is also getting a larger share of European Commission funds for building and refurbishing scientific facilities. In the next two years, he notes, Catalonia — mostly Barcelona — will get Pta 12 billion, compared with Pta 4 billion for Madrid, and Pta 5 billion for the Basque Country.

Miranda says this allocation will hurt Madrid because the city has so many more researchers and laboratories than Barcelona. The city has three times as many CSIC centres as Barcelona and twice as many universities.

But Federico Mayor, director of Madrid's Centre for Molecular Biology, says that both cities need to strengthen their scientific bases simultaneously. Jordi Mas, director of the Catalan Foundation for Research, agrees. "A strong Madrid and a weak Barcelona, or vice versa, won't help to position Spain in the future of European science," Mas says.

Andreu Mas-Colell, who runs the Catalan government's research department, says he would like Spain's science map to look like that of Germany, with competitive research distributed between many different cities. ■

## Primate centre promises insight into ape research

**Nina Schnapp, Munich**

A primate research centre with a difference opens this week in Leipzig. All of its work will be undertaken in full view of the public, who will be able to visit it to watch the four species of great ape housed there.

The Wolfgang Köhler Primate Research Centre, built at a cost of DM30 million (US\$13.5 million), is situated in the city's zoo. Its founders hope that the centre's 'open-door' approach will help to boost public confidence in their work.

"The new centre offers us a unique opportunity to compare the cognitive abilities of all four great ape species," says Michael Tomasello, its co-director. "We will use both observational and experimental methods to investigate how apes solve ecological and social problems."

Sixty great apes — chimpanzees, bonobos, gorillas and orang-utans — will be delivered to the centre from zoos across Europe. Some have arrived already. Their new home consists of indoor and outdoor enclosures designed to imitate the apes' natural surroundings. There are no bars — the apes are separated from visitors by moats or glass.

Research on intelligence, tool-using,

numerical skills and social behaviour will start in June, once the animals have settled down and formed social groups.

Visitors will be able to watch the scientists working with the apes through panes of one-way glass in hidden observation corners. This level of openness is important to the project, says Josep Call, co-director of the centre. "Seeing apes

solving problems in front of your eyes is very exciting — more so than seeing them on television."

Primate expert Jane Goodall, of the UK-based Jane Goodall Institute for Wildlife Research, Education and Conservation, has welcomed the centre. "Any new exhibit involving [highly endangered apes] that helps people to understand their true nature is to be applauded," she says. ■



Ape watch: the chimpanzee enclosure at the primate centre offers visitors unparalleled visual access.

MICHAEL SERES



# Ecologists score victory over controversial dyke project

David Cyranoski, Tokyo

A huge land reclamation project in south-western Japan is to be halted and partially reversed while researchers investigate its implications for marine biodiversity.

The move is seen as a significant victory for Japanese ecologists and seaweed farmers. The agriculture minister Yoshio Yatsu announced last week that the gates of a large dyke at Isahaya bay will be opened at some point while the study is done. Some ecologists believe that this will help the local marine ecosystem to recover.

Poor seaweed harvests and fishery catches led the agriculture ministry to ask an independent panel, chaired by Makoto Shimizu, a former University of Tokyo agriculturist, to look into the reclamation programme. The panel found that uncontrolled growth of diatom plankton was responsible for upsetting the ecosystem and depriving seaweed of nutrients.

Many ecologists blame the dyke, which was built in 1997 as part of a ¥249 billion (US\$2 billion) project to reclaim land around the bay for farming. The dyke created a reservoir inside the bay where sea water is purified into fresh water, a process that kills the clams and other shellfish that live in it. These shellfish would normally eat the plankton, which has since been free to grow out of control.

The panel recommends that the dyke's

gates be opened during a two-year study of the dyke's effects on seaweed growth. But it suggests delaying this action for up to a year to allow for safety measures to be taken against possible adverse outcomes such as flooding.

An agriculture ministry official says that it is still "considering other hypotheses besides the dyke" as the cause of the plankton growth. These include the impact of the nearby city of Fukuoka's diversion of water for drinking from a river that feeds the bay.

But some ecologists are critical of the delay in opening the gates. "There is no time to question which factor is most important — we must try to restore the ecosystem before it is too late," says Masanori Sato of Kagoshima University's environmental science department. ■



Flood of protests: the dyke at Isahaya bay has been blamed for damaging the local ecosystem.

## Network boost for southeast Europe

Quirin Schiermeier, Venice

The United Nations Educational, Scientific and Cultural Organization (UNESCO) is making a move to revitalize science in southeastern Europe. It plans to help build computer networks for research and education across the region.

A meeting on scientific cooperation in southeastern Europe, held in Venice last week, identified the lack of such networks as the largest single impediment to strengthening cooperation in the area.

Pierre Lasserre, director of UNESCO's European science and technology office, said the organization would provide US\$500,000 to help buy computers and network access in the region's less-advanced countries, including Bulgaria, Romania and the fragments of the former Yugoslavian federation. UNESCO called on the European Union to provide additional financial support.

The meeting, which was co-organized by UNESCO, the Academia Europaea and the

European Science Foundation, brought together representatives from 20 European countries, including ten in southeastern Europe. Those from the southeast complained that the area's science base suffers from a lack of support from its national governments.

The situation is particularly acute in the war-torn Balkans. "In the aftermath of the war, academic life has been starved out," said Nedžad Mulabegovic, rector of the University of Sarajevo in Bosnia-Herzegovina.

Speakers also reported a severe "brain drain" from the region. Almost half of Albania's scientists left the country during the 1990s, said Tamara Eftimi, rector of Tirana Polytechnical University.

The meeting agreed that the region's best prospects lay in better access to research facilities in western Europe, particularly for the training of its young scientists. "What we need is a European training space," said Lasserre. ■

## Indian rocket fizzles out as test launch fails to fly

K. S. Jayaraman, New Delhi

India's space science programme suffered a severe setback last week when the country's new launch rocket failed to take off on its maiden flight. The rocket, Geosynchronous Satellite Launch Vehicle (GSLV), cost 14 billion rupees (US\$300 million) to build and took ten years to develop.

Seconds after ignition, the GSLV's computer system aborted the 28 March take-off when one of the four strap-on motors in the rocket's first stage failed to develop sufficient thrust.

The aborted launch is a blow for the Indian Space Research Organisation (ISRO), which hopes that the project will end India's dependence on foreign launchers for its satellites. The ISRO also aims to gain a share of the global satellite launch business with the rocket.

These dreams are now on hold as several new systems in the GSLV, including an upper stage design imported from Russia, remain untested.

ISRO chairman Krishnaswami Kasturirangan said afterwards that failures were "normal" in the space business. He said he was happy that the safety system worked as designed and that the mission aborted before igniting the main solid booster. That would have destroyed the entire rocket and its payload, the 1,540-kilogram GSAT-1 telecommunications satellite.

One ISRO official said that the GSLV will be ready for testing again after a detailed examination of the strap-on motor. Another official said that the GSLV programme would be delayed by between six months and a year.

The failure comes at a bad time for the ISRO, which hoped that a successful GSLV launch would encourage government support for its plans for a moon mission and a satellite dedicated to astronomy. ■



Going nowhere fast: the test flight of India's GSLV rocket ended in disappointment.



## Companies join forces to crack the human proteome

**Washington** Ambitious plans to compile a catalogue of all human proteins and analyse their interactions were unveiled this week.

Myriad Genetics of Salt Lake City, Utah, is forming a \$185 million collaboration with software company Oracle and electronics corporation Hitachi to characterize the human proteome over the next three years. They hope to produce a proprietary database with "all human protein interactions, all biochemical pathways and a comprehensive catalogue of purified proteins".

Structural biologists will use data on protein structure to identify regions within a protein that are therapeutically important. The partners hope that this information will then play an important role in drug design.

## Europe faces up to fears of foot-and-mouth outbreak

**London** The number of cases of foot-and-mouth disease in Britain looked set to exceed 1,000 this week, as fears mounted that smaller outbreaks in nearby European countries may also prove difficult to control.

Dutch vets had confirmed ten cases by 2 April. The first seven were on farms relatively close to each other, but three subsequent outbreaks occurred on farms 30 kilometres away from the original cases. Dutch farmers now fear that the government programme of vaccination and slaughter may have failed to contain the virus.

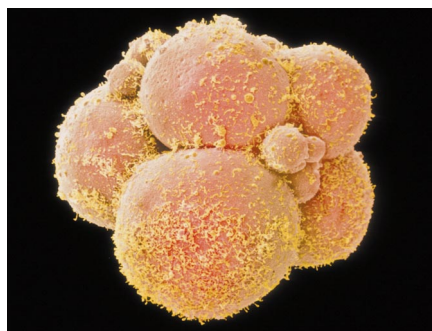
In Britain, pressure from farmers has forced Prime Minister Tony Blair to postpone local elections — and an expected general election — from 3 May. The number of cases is still increasing exponentially, as predicted by three epidemiological studies presented last week (see *Nature* 410, 501; 2001).

## Canadian panel backs stem-cell research

**Ottawa** Canadian scientists should have access to public money for research on stem cells and their derivation from human embryos and fetal tissue, an advisory panel told the Canadian Institutes of Health Research (CIHR) last week.

Lack of funding guidelines for stem-cell research led the CIHR — the country's main biomedical research agency — to set up a 10-member panel of scientists, bioethicists and policy experts to look into the issue. Janet Rossant, a developmental geneticist at the University of Toronto, is chairing the group.

The panel is asking scientists and members of the public to comment on its draft report, which recommends that the CIHR should fund research on stem cells derived from fetal



Medical potential: an eight-cell human embryo.

tissue and embryos left over from fertility treatments, but not from embryos created specifically for the purpose of providing stem cells. It proposes a moratorium on funding for work that involves mixing human stem cells with animal embryos.

The panel's final recommendations are due in June. Any move by the CIHR to fund stem-cell research might influence the debate in the United States, where some scientists fear that President George W. Bush will overturn a decision to allow limited funding for such work.

► <http://www.cihr.ca>

## Japanese research institute gets clean bill of health

**Tokyo** One of Japan's leading institutes for research into infectious disease is free to continue its work after a court rejected a claim that it posed a threat to public health.

When the National Institute of Infectious Diseases (NIID) relocated to central Tokyo in 1989, a court action was brought against it by local residents concerned that the pathogens studied by the institute could escape through the air-conditioning or in waste water. Teams of scientists from the United States and Britain were brought in by the NIID and the residents, respectively, to support their arguments during the trial.

A Tokyo court ruled last week that the scientific evidence did not support the residents' case. But the battle is not over — the residents have said they will appeal.

► <http://www.nih.go.jp/niid/index-e.html>

## France to address gender inequality

**Paris** The lack of women in top jobs in French science is to be investigated by the CNRS, France's national research agency.

Although more than 40% of CNRS staff are women, only a small number reach the most senior posts. Geneviève Berger, the agency's director-general, announced plans last week to set up a committee to research the causes of the gender inequality.

The CNRS committee complements the French research ministry's recent plans for a package of seven initiatives to promote

women in science, including a new 'women, science and technology' unit. This would aim to ensure that women get fair treatment when research policy is being implemented.

## AIDS vaccine gets off to a promising start

**San Francisco** An AIDS vaccine has shown promising signs in experiments with monkeys, researchers at pharmaceutical company Merck announced this week.

The vaccine consists of an inactivated cold virus into which the gene for the outer shell of HIV has been inserted. When injected into monkeys, the virus delivers the gene to the immune system, teaching it to recognize and attack HIV. Three monkeys that received the vaccine are still healthy eight months after being injected with a highly virulent form of HIV. Six out of eight control monkeys deprived of the vaccine have died.

The vaccine will not prevent HIV infection, but it might prolong the lives of people with HIV, without the need for drugs. Merck estimates that at least six years of clinical testing will be necessary before the vaccine is ready for widespread use.

## Double whammy for beleaguered lab

**London** The troubled UK drug-testing lab Huntingdon Life Sciences (HLS) suffered a double blow last week. Pharmaceutical company British Biotech announced it would terminate its remaining HLS contracts just days after HLS was forced off the main London Stock Exchange.

All companies on the London exchange must use at least two independent brokers to trade shares. But both of HLS's brokers stopped trading its shares last week after one, Winterflood Securities, suffered demonstrations outside its London offices.

British Biotech decided last June not to place new contracts with HLS, but recent pressure from activists led it to stop all HLS work at once. Protesters who demonstrated outside the company's Oxford office and harassed British Biotech staff now say they will target other HLS clients, such as Novartis, Roche and Bayer.



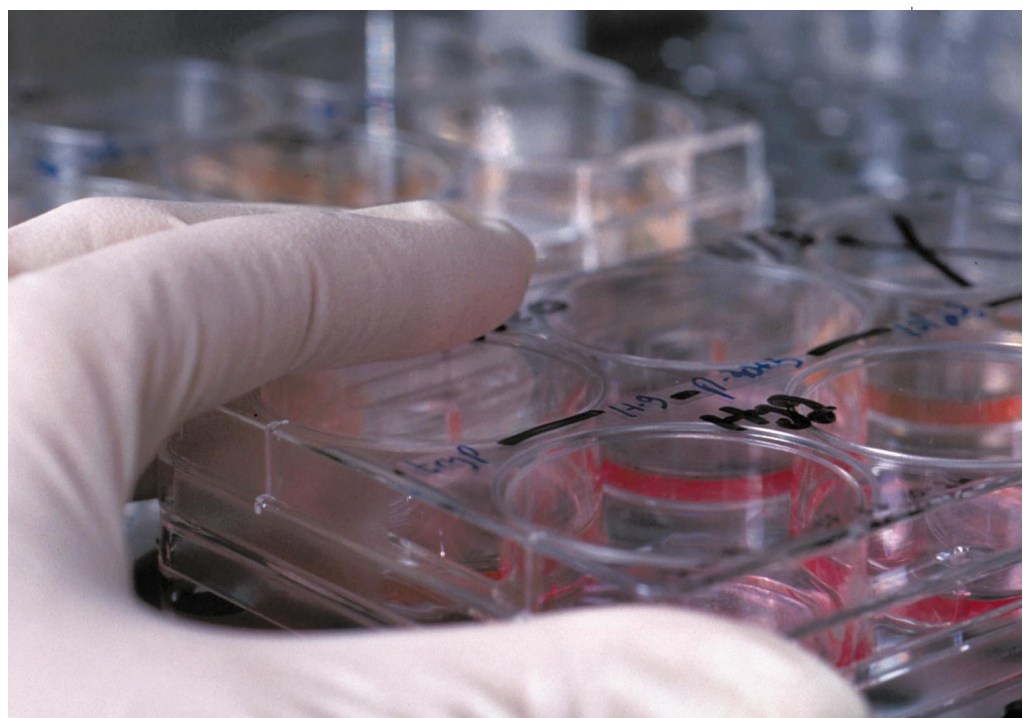
Dogged determination: continuing protests are causing significant problems for HLS.

# Can they rebuild us?

The idea of therapeutic cloning, which offers the potential of growing replacement tissues perfectly matched to their recipients, is falling from favour. But there are alternatives, as Peter Aldhous found out.

**T**ake two of the biological breakthroughs of the late 1990s and combine them to produce a medical miracle — that is the thinking behind therapeutic cloning. The achievements are the cloning technology that in February 1997 gave us Dolly the sheep<sup>1</sup>, and the successful creation the following year of cultures of human embryonic stem (ES) cells<sup>2</sup>. The promised miracle is the generation of ‘personalized’ replacement tissues to combat the ravages of ageing and disease. Genetically matched to the patient, these tissues would avoid the rejection problems that have always plagued transplant medicine.

ES cells come from blastocysts — tiny embryos, just a few days old, that consist of a hollow ball of cells. ES cells can develop into any type of cell, and so could be cultured to



Repair kit: can embryonic stem-cell cultures fulfil their promise of delivering replacement tissues?

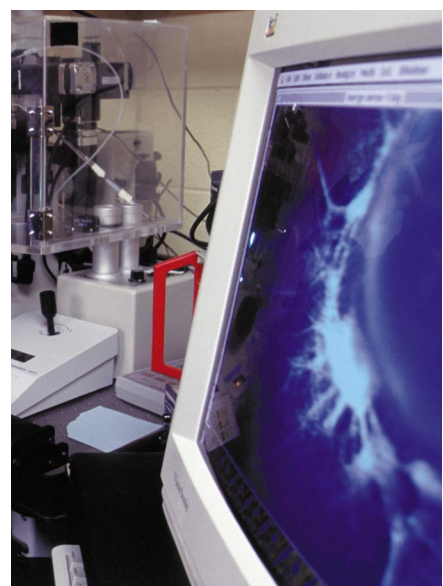
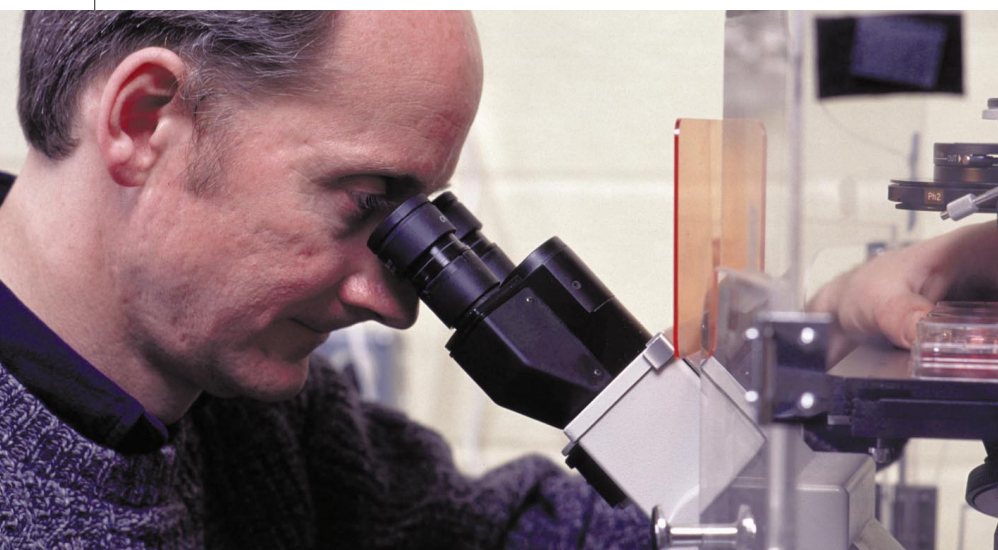
grow replacement tissues, such as cardiac muscle to graft onto a weakened heart. Therapeutic cloning aims to create ES cells that are genetically matched to the patient by using the technique that created Dolly. A healthy cell from a patient would be fused with a donor egg cell stripped of its chromosomes. This would produce an embryo which, given the right conditions, should develop into a blastocyst from which ES cells could be harvested.

## High hopes

Enthusiasm for therapeutic cloning was initially high. In a December 1999 article in *Nature*<sup>3</sup>, two leading cloning researchers declared their belief that such procedures

would bring “the greatest eventual benefit” from the technology. And over the past few years, therapeutic cloning has featured prominently in the popular press accounts.

So to the casual observer, it may come as a surprise that many experts do not now expect therapeutic cloning to have a large clinical impact. Aside from problems with the supply of human egg cells, and ethical objections to any therapy that requires the destruction of human embryos, many researchers have come to doubt whether therapeutic cloning will ever be efficient enough to be commercially viable. “It would be astronomically expensive,” says James Thomson of the University of Wisconsin in Madison, who led the team that first isolated



Cultural revolution: James Thomson was the first to isolate human embryonic stem cells.



ES cells from human blastocysts.

But the field of regenerative medicine is not in the doldrums — far from it. Some stem-cell biologists argue that it might be possible to treat patients by manipulating the ‘adult’ stem cells that reside in many of our tissues. Others are busy collaborating with immunologists to develop strategies that will allow tissues grown from ‘foreign’ stem cells — including ES cells — to evade the body’s immune system. And yet others believe that, in the long run, it may be possible to achieve the same goals as therapeutic cloning without a cloning step. They want to ‘reprogramme’ cells, reversing the developmental processes that made them adopt a particular specialized function, and turning them into ‘ES-like’ cells that can develop into any tissue.

### Double troubles

Therapeutic cloning is almost certainly possible. Researchers at Monash University and the company Stem Cell Sciences, both based near Melbourne in Australia, last year proved the principle by obtaining mouse ES cells from embryos that had been cloned from adult mouse cells<sup>4</sup>. But mammalian cloning is inefficient, even in the hands of the most skilled scientists. Of the 277 cells from Dolly’s ‘mother’ that were fused with donor egg cells, less than 30 developed to the blastocyst stage<sup>1</sup>. At the time, experts believed the efficiency would improve. But despite feverish efforts by groups worldwide, progress has been disappointing. “We don’t at the moment have any real handle on how to greatly increase the efficiency,” admits Alan Colman of PPL Therapeutics near Edinburgh, the company involved in the Dolly experiments.

For therapeutic cloning to become affordable, the cloning step would have to be conducted efficiently by technical staff at individual hospitals. Human eggs are also in short supply, and in high demand for *in vitro* fertilization procedures. Peter Mountford, chief scientific officer of Stem Cell Sciences, believes these problems can be overcome, and argues that it is too early to give up on therapeutic cloning — but his has become a minority view.

Although the progress in improving the efficiency of cloning has stalled, research with human ES cells has continued — albeit in a restricted number of labs (see ‘A tortured tale of supply and demand’, overleaf). They seem to proliferate well in culture<sup>5</sup>, and can develop in the laboratory into a wide range of different cell types<sup>6</sup>. At the Keystone Symposium on Pluripotent Stem Cells, held this February in Durango, Colorado, Melissa Carpenter of Geron, based in Menlo Park, California, reported on experiments in which she had allowed human ES cells to develop into ‘neural progenitor’ cells, which can develop into nerve cells. When she transplanted these into the brains of newborn rats, the cells seemed to

**T**oo much of this research is happening under the umbrella of biotech companies, which are understandably cagey.

continue their development and integrate into their new environment.

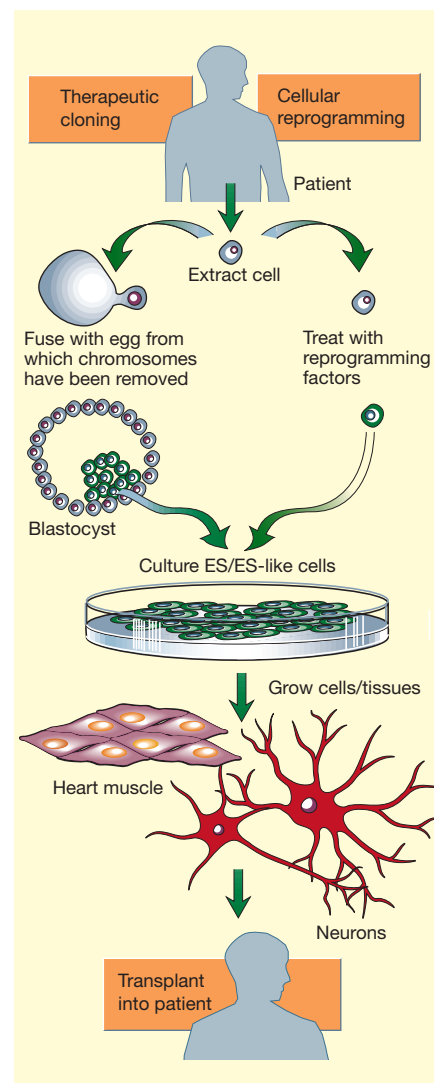
Given these advances, stem-cell biologists are cautiously optimistic about the prospects of growing replacement tissues from ES cells. Geron is pushing ahead, and hopes to move into clinical trials within five years. But if the ES cells do not come from embryos cloned from the patient’s own cells, the problem of rejection remains. In some cases, it may be possible to protect grafts grown from ES cells using relatively mild immunosuppressive drugs. The immune system has only restricted access to the brain, for example, so grafts of cells to replace the nerve cells lost in Parkinson’s disease might survive without much assistance. But in most tissues, grafts grown from foreign ES cells would quickly be rejected.

“The major issue in bringing this to reality is the immunological one,” says John Gearhart, a stem-cell biologist at Johns Hopkins University in Baltimore. And because immunosuppression renders transplant patients susceptible to infectious diseases and cancer, there is a big incentive to develop ‘tolerance’ strategies that would allow tissues grown from ES cells to escape the attentions of the immune system.

There are many ways in which this might be achieved. One idea is to use antibodies to block or disable receptors carried by immune cells involved in rejecting foreign tissues. In mice, temporary treatment with such antibodies around the time of a transplant seems to prevent rejection<sup>7</sup> — although such strategies risk rendering patients tolerant to any bacteria or viruses they encounter during the treatment. Maggie Dallman, an immunologist at Imperial College London, also warns that tolerance regimes that work in rodents often do not transfer so well to larger animals, or people. “That’s a general rule, and nobody quite understands why,” she says.

### Enter the engineers

Rather than manipulating the immune system to accept foreign grafts, some stem-cell biologists think it may be possible to genetically engineer ES cells so they become invisible to the immune system. “ES cells may need very little modification to make them universal donors,” speculates Alan Troun-



**Growth industry: the concept of reprogramming cells, rather than cloning them, offers an alternative route to generating ‘personalized’ tissue grafts.**

son of Monash University, a reproductive biologist who is now branching out into stem-cell research. Various strategies could be used. Dallman and her colleagues, for example, are investigating a protein called Notch, which helps regulate immune responses<sup>8</sup>. They suspect that stem cells could evade the immune system if they were engineered to produce a protein to which Notch binds.

Gearhart, meanwhile, is interested in the possibility of customizing ES cells by genetic engineering to make them match the intended graft recipient. The rejection of transplanted tissues depends heavily on proteins produced by genes within a chunk of chromosome 6 known as the major histocompatibility complex (MHC). Replace the MHC of ES cells with the patient’s MHC, and the immune system might be fooled into thinking the ES cells come from the patient. Replacing such a large gene sequence is technically difficult — but, argues Gearhart, not impossible.



Recent work with adult stem cells, however, has made some researchers question whether a tight focus on ES cells is necessary. Small numbers of stem cells exist in adult tissues, where they help to repair our bodies. Compared with ES cells, these adult versions are thought to have a more restricted capacity for development into different tissue types. But if they could be used as a source of replacement tissue, adult stem cells would avoid the destruction of a human embryo — a fundamental moral objection to approaches based on ES cells.

Adult stem cells could be harvested from healthy donors and used to grow replacement tissues for patients needing grafts. Once again immunosuppression or tolerance would probably be needed, but a novel way of using adult stem cells might provide a suitable tolerance strategy. Bone marrow contains haematopoietic stem cells (HSCs), which give rise to all of our blood cells, including those of the immune system. When HSCs are transplanted into the bone marrow of the recipient, the immune system can enter a 'chimaeric' state in which some of its cells are derived from the transplanted HSCs. These would, in theory, prevent the immune system from reacting against other cells transplanted from the same donor<sup>9</sup>. Last year, for instance, Judith Shizuru, Irving Weissman and their colleagues at Stanford University in California showed that mice given transplants of highly purified HSCs subsequently accepted heart grafts from mice genetically identical to those from which the HSCs came<sup>10</sup>.

But if a patient's own stem cells could be used to grow replacement tissues, there would be no need to worry about rejection. With this aim in mind, researchers are again looking to bone marrow to provide a solution. Bone marrow contains stem cells that can give rise to a range of tissues including bone, cartilage and muscle. In April 1999, researchers with the company Osiris Therapeutics in Baltimore showed that cultures of these cells retain this potential<sup>11</sup>. And in this issue of *Nature*<sup>12</sup>, Piero Anversa of the New York Medical College in Valhalla and his colleagues describe experiments in which they injected stem cells from mouse bone marrow directly into the cardiac muscles of mice with damaged hearts. They found that the stem cells developed into muscle cells and blood vessels, helping to repair areas of dead tissue. These experiments raise the possibility of repairing a patient's failing heart with cardiac muscle grown from his or her own bone marrow stem cells<sup>13</sup>.

### Career change

Recent experiments in mice have also revealed that adult stem cells can develop in entirely unexpected ways. Neural stem cells from the brain, for example, have been transplanted into bone marrow, where they

## A tortured tale of supply and demand

Given the breadth of their potential, one might expect that human embryonic stem (ES) cells would be the focus of attention for hundreds of research groups. But so far, only a dozen or so teams have entered the field. The issue was initially one of a shortage of cells. But now the main problems are political — with fears that the new US administration will ban federal funding for ES-cell research looming large.

So far, the main distributor of human ES cells has been the WiCell Research Institute, a non-profit spin-off from the University of Wisconsin in Madison, where in 1998 the cells were first cultured in James Thomson's lab<sup>2</sup>. His work was funded by the company Geron of Menlo Park, California, which has certain exclusive commercial rights to develop the ES cells for therapeutic applications.

In February last year, the University of Wisconsin announced that WiCell would soon start making Thomson's ES cell lines available to other research groups. Some researchers were initially concerned that WiCell wanted wide-ranging rights to rescind permission to work on the cells and to demand that they be destroyed. Those rights have since been restricted, and will only be enforced under specific

circumstances — for instance if researchers use the cells for additional projects without written permission.

Given the time that has elapsed between the cells' creation and WiCell's formation, say stem-cell researchers, Geron got an important head start. Although researchers who use WiCell's ES cell lines can patent discoveries made using the cells, they may find that much of this territory has already been staked out. "We have submitted 36 patent filings on these cells," says David Greenwood, Geron's senior vice-president for corporate development.

Other supplies are now available. A group headed by Alan Trounson at Monash University, near Melbourne in Australia, is responding to requests to obtain cells from its human ES cell lines<sup>6</sup> — although high demand is putting the lab under pressure. "We are short on staff," says Trounson. Other ES cell lines are soon expected to become available from the Rambam Medical Center in Haifa, Israel.

But increased availability does not mean open access. Current legislation in France and Germany, for instance, prohibits embryo research, including work on ES cells — although the French government is proposing to lift its ban. In the United States,

meanwhile, uncertainty as to whether federal funds will be freed to support ES-cell research is hampering progress.

Last summer, the previous US administration concocted a compromise that would allow researchers to use federal funds to work on ES cells, provided the ethically contentious step of isolating the cells from a human embryo had been achieved using other funding sources. But President George W. Bush's administration may now block the use of federal funds for ES cell research. And in the current state of limbo, few scientists have responded to a call from the National Institutes of Health (NIH) for ES-cell research proposals. By the 15 March deadline for documents to show that proposed research will comply with NIH guidelines, just three submissions had been received.

Even if federal funding is released, there may still be problems. The NIH has published criteria — including standards for informed consent from 'parents' of the embryos from which the cells were harvested — with which suppliers of ES cells to federally funded researchers must comply. WiCell's current cell lines do not meet these standards, and the NIH is still reviewing compliance documents submitted from Monash. **Joanna Downer**

developed into blood cells<sup>14</sup>. Bone marrow stem cells have also been shown to migrate to the brain after being injected into the bloodstream, where they develop into cells that appear to be neurons<sup>15,16</sup>. These experiments have fuelled hopes of treating patients with their own adult stem cells.

But even the enthusiasts accept that there is a long haul ahead before therapies based on these discoveries are ready for the clinic. "We need to make this more robust," says Helen Blau, who works on adult stem cells at Stanford. Showing that small numbers of stem cells can migrate to another site in the body and develop into a cell type appropriate for that tissue is one thing; using them to repair damaged or diseased tissues is another.

Improving the situation will entail a search for cell-surface markers to identify the stem cells that can transform into a wide range of tissue types, and the development of

methods to purify and selectively culture them. It may also require the discovery of the biochemical signals that attract stem cells to sites of tissue damage and direct their development to effect a repair.

Given these obstacles, some researchers believe it is also worth taking on the field's toughest challenge — finding a way to reprogramme any of the body's cells to create ES-like cells matched to the intended recipient without cloning an embryo. Interest stems in part from experiments reported in 1997, in which researchers led by Azim Surani of the Wellcome/CRC Institute of Cancer and Developmental Biology in Cambridge fused mouse white blood cells with embryonic germ cells<sup>17</sup> — cells from the developing reproductive system that share many characteristics of ES cells. The white blood cell nuclei appeared to return to an embryonic state.

Researchers at several of the leading com-

panies interested in regenerative medicine are now rumoured to be stripping the nuclei from ES cells and embryonic germ cells, and fusing them with various types of cells in an attempt to wind back the developmental clocks of the latter. In the process, they hope to learn how to rewind cell development without using ES cells.

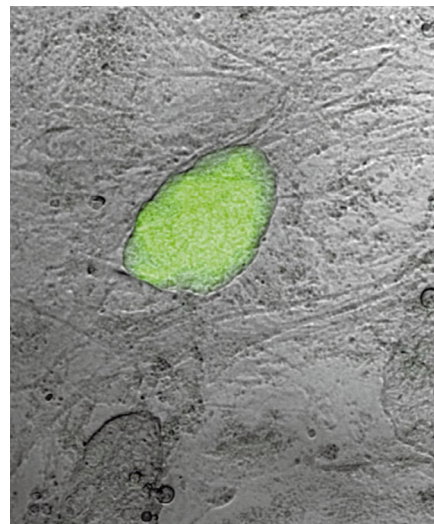
One such company, PPL Therapeutics, claimed in January to have reprogrammed skin cells to form ES-like cells, some of which developed into heart muscle cells. But PPL has annoyed other researchers by refusing to produce data to back up the claim. The company also will not confirm whether cell fusion, or some other technique, was involved. "I don't think this is helpful," says Rudolf Jaenisch of the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology. "If you make an announcement, you have to say how you did it."

Indeed, the commercial secrecy cloaking much of the work on cellular developmental reprogramming is causing widespread frustration. "Too much of this is happening under the umbrella of biotech companies, which are understandably cagey," says Richard Gardner of the University of Oxford, who last year chaired an expert panel that reported on the issues surrounding stem-cell research for Britain's Royal Society.

Other approaches thought to be under investigation behind closed company doors may be inspired by work on the African clawed toad, *Xenopus laevis*. Cloning amphibians is technically less challenging than cloning mammals — *Xenopus* have been routinely cloned since the 1960s. Biologists have recently started to identify the molecules in *Xenopus* egg cells that underpin the developmental repro-



Fresh start: Azim Surani's cell fusions (right) suggest that developmental reprogramming is possible.



gramming involved. And these findings provide hints about some of the changes needed to wind back development without a cloning step.

The DNA in every cell is associated with proteins that regulate the expression of the cell's genes. As cells move towards their specialized adult functions, some of these proteins are removed, while many more are added. So, among other processes, reprogramming means undoing these changes. In the 1990s, while studying cloning in *Xenopus* at the National Institute of Child Health and Human Development in Bethesda, Maryland, Alan Wolffe implicated a protein called nucleoplasmin in this process. Nucleoplasmin helps strip DNA from the histone proteins around which it is wound in the chromosomes of mature cells<sup>18</sup> — thought to be a necessary stage in cellular reprogramming.

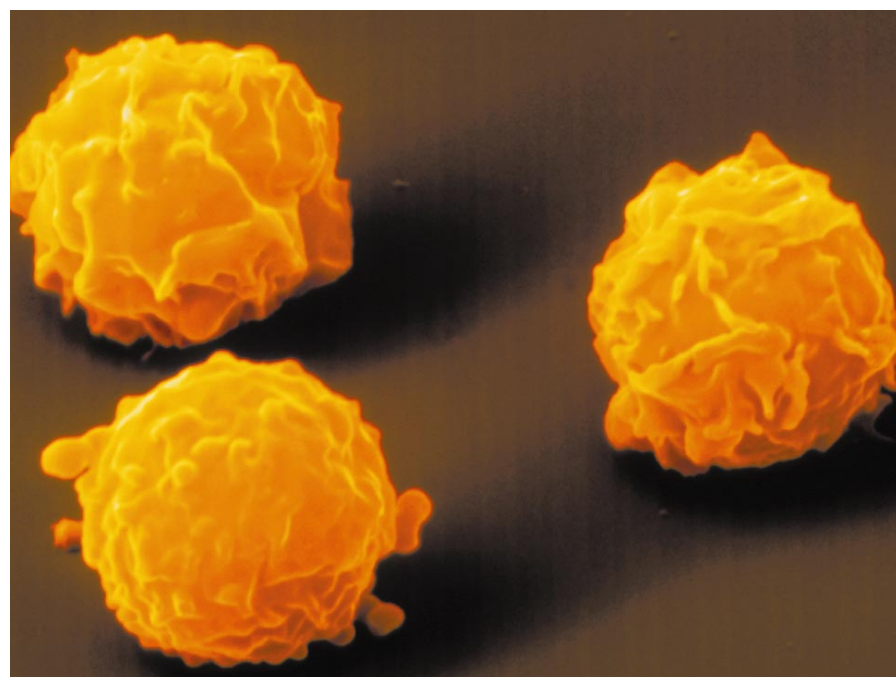
And last year, he led a team that showed that an enzyme called ISWI removes another protein, called TATA-binding protein, that associates with the DNA of adult cells<sup>19</sup>.

Such discoveries may merely be scratching the surface of the reprogramming mechanism. But other researchers intend to conduct systematic screens for cellular reprogramming factors. Surani, for instance, is now embarking on experiments with cells engineered to contain 'reporter' genes that should switch on if the cells are reprogrammed. Surani plans to insert a library of genes into the engineered cells, to find those that activate the reporter gene.

Enthusiasm for therapeutic cloning may have dimmed, but regenerative medicine is still a hotbed of activity, with molecular and cell biologists, immunologists, geneticists and developmental biologists all claiming a piece of the action. If the pace of discovery holds up, stem cells might eventually deliver a medical miracle. ■

**Peter Aldhous is Nature's chief News and Features editor.**

1. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385**, 810–813 (1997).
2. Thomson, J. A. *et al. Science* **282**, 1145–1147 (1998).
3. Gurdon, J. B. & Colman, A. *Nature* **402**, 743–746 (1999).
4. Munsie, M. *et al. Curr. Biol.* **10**, 989–992 (2000).
5. Amit, A. *et al. Dev. Biol.* **227**, 271–278 (2000).
6. Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A. & Bongso, A. *Nature Biotech.* **18**, 399–404 (2000).
7. Honey, K., Cobbold, S. P. & Waldmann, H. *Immunol. Res.* **20**, 1–14 (1999).
8. Hoyne, G. F. *et al. Int. Immunol.* **12**, 177–185 (2000).
9. Wekerle, T. & Sykes, M. *Annu. Rev. Med.* **52**, 353–370 (2001).
10. Shizuru, J. A., Weissman, I. L., Kernoff, R., Masek, M. & Scheffold, Y. C. *Proc. Natl Acad. Sci. USA* **97**, 9555–9560 (2000).
11. Pittenger, M. F. *et al. Science* **284**, 143–147 (1999).
12. Orlic, D. *et al. Nature* **410**, 701–705 (2001).
13. Sussman, M. *Nature* **410**, 640–641 (2001).
14. Bjornson, C. R., Ritz, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. *Science* **283**, 534–537 (1999).
15. Brazelton, T. R., Rossi, F. M., Keshet, G. I. & Blau, H. M. *Science* **290**, 1775–1779 (2000).
16. Mezey, E., Chandross, K. J., Harta, G., Maki, R. A. & McKercher, S. R. *Science* **290**, 1779–1782 (2000).
17. Tada, M., Tada, T., Lefebvre, L., Barton, S. C. & Surani, M. A. *EMBO J.* **16**, 6510–6520 (1997).
18. Dimitrov, S. & Wolffe, A. P. *EMBO J.* **15**, 5897–5906 (1996).
19. Kikyo, N., Wade, P. A., Guschin, D., Ge, H. & Wolffe, A. P. *Science* **289**, 2360–2366 (2000).



Rich resource: stem cells from bone marrow could be used to repair a variety of tissues.



# Breaking the nuclear taboo

Despite public fears, nuclear-powered spacecraft are back on the agenda. Tony Reichhardt looks into the latest plans for planetary exploration.

Studying the fringes of our Solar System can be a frustrating business. Jupiter's moon Europa, with its ice-covered ocean, and Saturn's cloud-shrouded Titan are exciting places. The problem is getting there. Not only are the outer planets far away, but extra fuel is needed to brake a spacecraft into orbit around a tiny moon circling a massive planet. The exorbitant costs of launching spacecraft weighed down with sufficient fuel rule out most missions.

Delegates at a NASA meeting on new approaches to exploring the outer Solar System, held in February at the Lunar and Planetary Institute in Houston, kept coming back to an old but controversial solution: nuclear-powered rockets. Giovanni Bignami, scientific director of the Italian space agency ASI, puts it bluntly: "You can't seriously go to the planets unless you use nuclear propulsion. All the rest is junk." But funding agencies, sensitive to public opinion, will need convincing. Protests against the small amounts of onboard plutonium at the launch of NASA's Galileo mission to Jupiter in 1989 and the Cassini Saturn mission in 1997 have turned 'nuclear' into a forbidden word.

Rockets produce thrust by forcing a high-pressure gas to escape through a nozzle,

which accelerates the craft in the opposite direction. The space shuttle's main engines, for example, produce high-pressure gas by combining liquid oxygen and hydrogen. A nuclear thermal rocket would use the heat generated by a small nuclear reactor to turn liquid hydrogen into high-pressure hydrogen gas. Nuclear rockets have the edge because, in weight terms, they use less fuel to produce comparable thrust.

Billions of dollars have been spent on nuclear propulsion research since the 1960s, but an engine has yet to be tested in space. Interest flared up in the early 1990s, when NASA was asked to consider sending astronauts to Mars. But NASA's proposals were too expensive, and when the programme stalled, so did interest in nuclear rockets.

Now, for the first time in years, propulsion experts find themselves with modest funding to study engines powered by nuclear fission. Researchers at NASA's Glenn Research Center in Cleveland, Ohio, believe a 500-kilogram nuclear-powered spacecraft could be flying past Pluto in as little as six and a half years after its launch — around four years faster than conventional rockets could manage. Glenn's veteran nuclear rocket designer Stan Borowski puts costs for the ten-year development of a nuclear thermal engine at \$1.5 billion.

In Italy, Bignami's agency is funding preliminary work on a concept developed by Nobel prizewinning physicist Carlo Rubbia, now at the University of Pavia. Rubbia's proposal would use americium-242 rather than uranium for fuel. Conventional nuclear thermal engines generate heat from fission, and then transfer the heat to a separate container of hydrogen. Rubbia proposes placing the americium in direct contact with the hydrogen, making the heat transfer more efficient (*Nature* 397, 374; 1999). Work on 'Project 242' is currently focused on bonding americium to a substrate.

Another possibility is nuclear electric propulsion (NEP), in which electricity generated by a reactor is used to expel a stream of ions from the back of the rocket. The thrust is small, but even a small thrust will eventually accelerate a spacecraft to high speeds in the near vacuum of space. Ground tests of the Safe Affordable Fission Engine (SAFE, in case anyone misses the point), a NEP engine developed at NASA's Marshall Space Flight Center in Alabama, are already under way.



Blasted: the small amounts of plutonium used in Cassini's Saturn mission provoked protests.

Ground tests are one thing, but building an engine and flying it in space is another. So will NASA make a serious go of nuclear propulsion this time? US missions to the outer planets are currently in a confused state. NASA wants to launch a Pluto mission in the next few years — proposals for the project are due in this week and one group is known to be submitting a nuclear propulsion option. But the new US administration is keen for NASA to prioritize work on propulsion technologies. NASA science chief Ed Weiler has said the propulsion research will "look at what's really needed" to make the Pluto mission successful, including nuclear thermal rockets and NEP.

Looming over discussions at last month's meeting was the question of how the public will react. Borowski's group at Glenn is already preparing a public information campaign. Others hope that President George W. Bush will provide the backing that his predecessor, Bill Clinton, declined to give.

All of which creates a cautious sense that perhaps, finally, nuclear propulsion's time has come. For some, it clearly feels like now or never. As William Jeffrey of the US Defense Advanced Research Projects Agency said at the NASA meeting: "If we can't make this a compelling argument now, we probably can't in our lifetime."

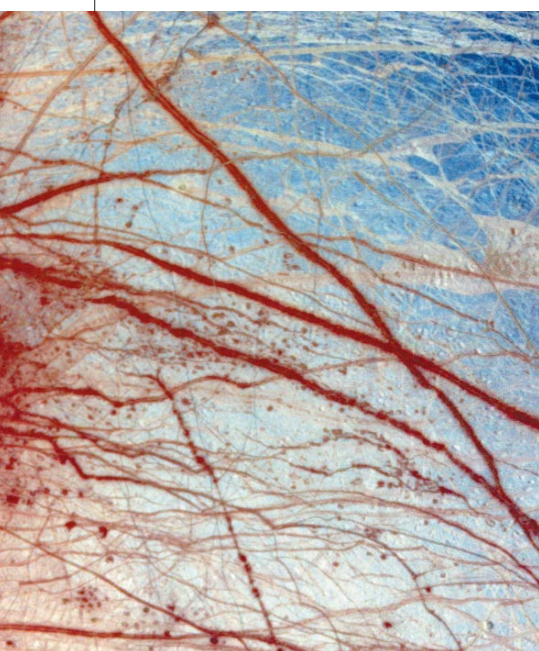
Tony Reichhardt is a contributing correspondent for *Nature*.

Project 242

♦ [http://ars.rm.asi.it/~webars/Project242/P242\\_main.html](http://ars.rm.asi.it/~webars/Project242/P242_main.html)

Safe Affordable Fission Engine

♦ <http://stday.msfc.nasa.gov/presentations/IST3.pdf>



Out of reach? The weight of fuel needed makes it expensive to put a craft in Europa's orbit.

## EU will gain from funding eastern European centres of excellence

*Sir*—The differences between standards of research, defence and industry in European Union (EU) member states and other central or eastern European countries are undeniable. These 'gaps' have led some people to argue that, if the latter countries join the EU, the lack of cohesion between them and the existing members could destroy the whole structure. Although this point of view could be challenged, we propose a practical solution: the systematic upgrading of research facilities in central and eastern European countries.

Before 1989, central and eastern European countries could not contribute properly to the development of advanced research equipment and devices going on in the West. Eleven years after the fall of the Berlin wall, the former communist countries have made only a little headway in commercial production and purchase of high-tech scientific equipment. Even in the countries that are making most rapid progress and have developed the skills to make non-serial high-tech equipment, the main obstacle to attaining excellence in experimental research remains either the insufficient performance of research facilities, or the lack of them altogether. This, together with low salaries, reduces the appeal of institutes in the region.

There are two main problems with upgrading. One is that only EU member states have access to the structural funds dedicated to 'cohesion building'. The other is that many experts and decision-makers do not consider that this type of assistance should cover quality of scientific research. They fear that money will go to unqualified people just because they are in a poor country, and the funding will still be unlikely to bring them up to a useful standard. Yet this presumed incompatibility between cohesion and excellence is more apparent than real: Margaret Sharp has pointed out (*Research Policy* 27, 569; 1998) that the concept of excellence puts emphasis on the EU as a whole, and cohesion is about helping poorer countries and regions to catch up. This is a pertinent analysis, identifying the two concepts as complementary. The next step, of course, is specific action to achieve cohesion.

One way of dealing with the issue is the recent action of the European Commission in identifying centres of excellence in the pre-accession countries. It seems quite likely that the majority of the 34 centres could receive long-term funding to establish them as European or regional attractors for research and training

activities, and as poles of local economic development. Such funds should be used mostly for upgrading research facilities. The European Council and the European Commission now need to stimulate the growth and consolidation of the selected centres via the structural funds of the EU.

The upgrade of the most promising research facilities requires a coordination mechanism: for example, financing could be shared along the lines of one national or regional euro for every European one. If some centres of excellence could serve several neighbouring countries, consortia could be formed between countries or regions to share the cost.

**Simeon Anguelov (former Bulgarian ambassador to France)\*, Norbert Kroo†, Pierre Lasserre‡, Pierre Papon§**

\*79 Quai André Citroën, Paris 75015, France

†Hungarian Academy of Sciences, Roosevelt sq. 9, H-1051 Budapest, Hungary

‡UNESCO Venice Office-Regional Office for S&T in Europe, 1262/A Dorsoduro, 30123-Venice, Italy

§Ecole supérieure de physique et chimie industrielles, 10 rue Vauquelin, 75005 Paris, France

## Spain is a closed culture to foreign researchers

*Sir*—We wish to add a foreigner's perspective on the current debate (for example, see *Nature* 407, 659; 2000 and 410, 14; 2001) about the Spanish research environment, in which we have each spent more than five years.

In a policy at odds with the rest of the EU, posts for non-EU nationals are simply not available in Spain's huge public sector. As for EU nationals, those with a degree from a country where the state does not regulate the profession of lecturer or researcher cannot apply for serious posts in Spain without either having three years' experience in their country of study or 'homologating' their degrees—having them approved by the Spanish authorities.

The homologation procedure is both arbitrary and extremely inefficient. Degrees from EU and other countries are homologated only after years of administrative stonewalling, if at all. One of us, for example, has published more than 100 papers in international publications, yet spent four years trying unsuccessfully to get homologation for an honours degree in mathematics from the University of Cambridge. This example is illustrative. The tiny number of foreign researchers in Spain compared with the rest of the EU suggests lack of compliance with EU directives on free movement of labour.

The issues of consistent government underinvestment and of closed-shop practices—due in large part to an

uninhibited culture of nepotism and patronage ill-befitting a democratic, developed country—have already been raised in the Opinion and Correspondence referred to above and in other articles. But the way these issues affect Spanish research merits further clarification.

Government underinvestment leads to many researchers spending 10 or more years on poorly paid, short-term (one year or less) contracts or grants, just to reach the starting salary of a public-sector school teacher. About half the lecturers in Spain are '*asociados*' on one-year contracts; many full-time *asociados* earn little more than 14,000 euros (US\$12,900) a year. Generally speaking, only short-term contracts and grants are available to foreigners, and these are mainly restricted to EU nationals.

Closed-shop practices can lead to posts being treated as largesse and PhDs becoming little more than probationary employment periods, lasting for six years or even more. This encourages a drift towards little or no supervision, little research of any value, and the award of a PhD becoming a formality for those who go the distance. On the way, PhD students, and others on short-term contracts, may have to work on all manner of tasks, a situation exacerbated by slack financial controls. Not surprisingly, a large number of PhDs are abandoned. In this climate, foreigners who come to study for a PhD may end up going home without one.

Such a deep-rooted endogamic culture might be broken only by drastic measures such as prohibiting the move from a doctorate to a post in the same institution and requiring a minimum number of international publications for higher-level posts. Although it is important to entice Spanish researchers working overseas to return (currently it is very difficult for them to do this), focusing on this issue—instead of on the issue of researcher mobility in general—draws attention away from the root of the problem, that of impermeability to external candidates.

It is ironic that the Spanish government is pressing for more of this mobility in the context of the next EU research programme, when it has so far made little attempt to put its own house in order. It is time the many dedicated, hard-working and talented researchers in Spain were provided with a system and a budget under which they can produce quality results.

**Simon Pickin**

IRISA, Campus de Beaulieu, 35042 Rennes, France

Other signatories to this letter:

Peter Breuer, Carlos III University, Spain

Gennady Fikshan, Nortel Networks, Spain

Salvador López Mendoza, Autonomous National University of Mexico

Amparo López Gaona, Autonomous National University of Mexico

Carlos Sánchez Tarnawiecki, Fundacio Bosch i Gimpera, Peru



# Cultures of knowledge

Europe can benefit from East–West collaboration rather than globalization.

## Wolf Lepenies

The expansion process of the European Union has a focal point: the so-called *acquis communautaire*, an inventory of what has already been achieved by the union. This is an arrogant concept, especially in science policy. What matters far more is an *acquis commun*, the common basis of a spiritual and intellectual Europe. This is what we must reflect on if we wish to attain the “civilization of solidarity” that Walther Rathenau once defined as the goal of a European unification process.

I have more than ten years’ experience of institution building in central and eastern Europe. My own institution, the Wissenschaftskolleg zu Berlin, has been involved in building or supporting the Collegium Budapest, the New Europe College in Bucharest, the Bibliotheca Classica in St Petersburg and the graduate school for social sciences in Warsaw. We are currently setting up ‘Blue Bird’, an agenda for civil society in southeastern Europe, in Sofia, and are discussing whether it is feasible to create a centre for the social sciences and humanities in Tiflis, where scholars from the southern Caucasus countries would work together.

These institutions are in different stages of their development: the Collegium Budapest celebrates its tenth anniversary this year; the New Europe College is now turning into a regional centre; the Bibliotheca Classica, a research library for classical languages affiliated with a high school, has yet to attain financial security and to find an appropriate building; our support for the graduate school in Warsaw has come to an end. It is too early to predict what will happen in Sofia and Tiflis.

My experience is limited to building and supporting institutions associated with the term ‘institute for advanced study’ — where scholars from different disciplines and nationalities spend time together to pursue, either individually or in groups, their own research projects away from their usual university duties. I am not going to discuss here the enormous political, financial and intellectual difficulties that we had to overcome to realize these projects. Rather, I will mention some insights and conclusions resulting from our institution building.

The sudden liberation of the East in 1989

quickly exposed the gulf between the rhetoric and the political/cultural reality of Europe. We in West Germany, for instance, had allegedly been longing for our brothers and sisters in the East for more than 40 years: yet after unification, we were like the English poet Charles Lamb, who tried all his life to like Scotsmen and was finally obliged to desist in despair. Convinced by our own capacity and driven by unlimited benevolence, we saw ourselves as ‘experts beyond experience’. In truth, we were ‘experts before experience’, and quickly discovered that the combination of a splendid idea and a lot of money is not irresistible.

## Divided policy

Even today, foreign science policy in the East suffers from political cleavages that are much deeper than in the West. If I were asked by colleagues in the East for a single piece of advice, I would refer them to T. S. Eliot’s observation, that the survival of a parliamentary system requires constant dining with the opposition.

Another lesson we learned on arrival in the East was to realize that the division of

Europe was not just a problem for them — it was also, although we were largely unaware of it, a problem for us. Being cut off from intellectual and cultural traditions that made cities such as Budapest, Warsaw and Bucharest centres of great intellectual attraction had impoverished the West just as their isolation had impoverished the East. By helping to recover and reconstruct great European traditions of learning and scholarship, we were helping ourselves. In Budapest, for instance, we were provided with a fresh insight into the Islamic legacy of Europe; in Bucharest we learned that the Orthodox world view and a Byzantine past are also a part of our intellectual present; in St Petersburg we became aware that Latin and Ancient Greek are not dead languages but that they had played a vital part in preserving great European traditions of learning and of scholarship under dictatorial rule.

When we first came to the East — not as visitors with a limited visa but as friends who might stay as long as we wished — we were more impressed by richness than by poverty. Institutions were poor, instruments were lacking, but in a few cases intellectual traditions had prevailed and debates were conducted with an earnestness and a responsibility that were largely unfamiliar to us. This experience of finding treasure where we had been told that there would be only trash has remained with me ever since. We wanted, therefore, to aim at strengthening local cultures of knowledge, rather than merely

Lessons from the past: the West was as impoverished as the East by the isolation of cities such as Warsaw (left) and Budapest (below).





**Educational trip: contact with St Petersburg revealed that Latin was still flourishing.**

▶ exporting our own. We wanted to build institutions that would motivate the best local brains to stay home as well as attracting the best scholars from abroad.

We were reminded quickly, and sometimes brutally, that nationalism still exists in science and in scholarship. So, for instance, we did not create a Collegium Hungaricum in Budapest, and we tried to make sure that neither the New Europe College in Bucharest nor the graduate school in Warsaw are seen as Romanian or Polish institutions, respectively. The Blue Bird initiative in Sofia is not a Bulgarian enterprise, but a joint venture of Balkan neighbours, and we will give up our efforts in Tiflis if it turns out that the institute we want to create there is unable to serve scholars from all three countries in the southern Caucasus.

It has also been our deliberate policy to avoid any bilateral arrangement. Six European countries are collaborating in Budapest; Switzerland and Germany are cooperating in Bucharest; France and Germany are joining forces in Warsaw. Extending the possibilities and chances of fundraising was one motivation for this policy, but also cultural cooperation very often leads to attempts at cultural domination — multilateral arrangements are an effective check against such attempts.

European-US cooperation has been important, first by creating the New Europe Prize (US\$50,000 each given to eight institutions from 1993 to 1997) for newly emerging institutes in central and eastern Europe. The preservation of European cultural and intellectual diversity has some roots in emigration to the United States. To a much larger degree than hitherto, European-US joint ventures must become an important element of the European landscape.

The multilateral coalitions I have men-

## A common communist past has not made friends of those who were enemies for long periods of their history.

tioned consist, as a rule, of private foundations and of government agencies. This public-private mix has a considerable political impact by showing that society and the state are not the same and that their cooperation is indispensable for the functioning of modern democracy. Multilateral arrangements also demonstrate the advantages of intellectual diversity and remind us that Europe consists of an assembly of translated cultures.

One of our greatest illusions in 1989 was the assumption that western-style institutions in an eastern country would be attractive for their immediate neighbours. Almost the contrary was true: people wanted to visit institutions in the 'real' West rather than institutions they considered substitutes. One of our greatest success stories was when a young political scientist from Bratislava accepted an invitation from the Collegium Budapest rather than a fellowship at a very well-known institute for advanced study in the West.

We also had to learn that a common communist past has not made friends of those who were enemies for long periods of their history. Yet we have not given up the idea that the institutions we have helped to create or to survive should strengthen their idiosyncratic capacities by cooperating in a regional informal network.

The experience of the Collegium Budapest, where six European countries are cooperating, has led us to invite other European countries to join the Wissenschaftskolleg zu Berlin. Members must be on the board, and thereby help determine our intellectual policy, and also contribute substantial funds. Switzerland has just become an institutional member, and at least two other European countries will join us before long. 'Europeanization' of national institutions of higher learning might become an interesting element of European science policy in the future, being part of a European public sphere below the Brussels level and above national levels of communication.

### Politeness and science policy

In the past decade, we have witnessed the transition towards the market, democracy and the rule of law in the East and, at the same time, the dramatic change of political structures, legal regimes and economic paradigms in the West. These revolutionary developments highlight the fact that the key problems of all industrial societies require long-term policies for solutions. Ecological constraints will force us to develop a new contract between man and nature; growing

structural unemployment will require re-evaluation of the traditional notion of work as central to society; mass migration demands a much greater cultural flexibility in our daily life and in our institutional arrangements than we have developed so far. Science and technology, too, have entered the age of uncertainty, and we scholars are not well prepared for the challenges lying ahead.

Politeness in science policy is a willingness to learn that prevails over the inclination to be satisfied with what one already knows. The latter inclination usually carries with it the impulse to teach others — whether or not they wish to be taught. In Europe especially, we have well-established teaching cultures (Belehrungskulturen); politeness requires that we become learning cultures (Lernkulturen) again. This would demonstrate that the wish to understand prevails over the inclination to judge.

We learned very quickly after 1989 that two attitudes are most detrimental to effective cooperation between West and East in Europe: first, obviously, is meanness; the second is demonstrative charity, which has a counterproductive air of condescension. A more productive course is to acknowledge mutual need.

Politeness is the willingness to pay attention to context, detail, and minute differences in the attitudes and styles of everyone involved in a transaction. Regional competences have to be acquired to do justice to local knowledge: more than ever before, and in contrast to the detrimental rhetoric of globalization, we must strengthen the importance of local cultures of knowledge.

We in the West should take the recent historical developments on our continent as a most welcome opportunity to learn from each other, instead of simply using it as a pretext to teach others. Such a view might help to preserve the diversity of cultural and intellectual orientations that constitute the special richness of Europe.

Vice-president Dan Quayle was absolutely correct when he predicted, immediately after the overturn of the communist regimes on the 'old' continent, "The recent developments in Eastern Europe are irreversible. But this may change any minute." Never before has a linguistic blunder turned so quickly into political prophecy. In science policy, we should try not to fall behind that accidentally expressed wisdom. ■

*Wolf Lepenies is rector of the Wissenschaftskolleg zu Berlin, the Institute for Advanced Study in Berlin, Wallotstrasse 19, 14193 Berlin, Germany.*



# Playing the numbers game

How the cost of a can of Coke could help solve the world's population crisis.

## **Beyond Six Billion: Forecasting the World's Population**

edited by John Bongaarts  
& Rodolfo A. Bulatao  
*National Academy Press: 2000. 258 pp.*  
\$29.95, £21.95

## **Population and Climate Change**

by Brian C. O'Neill, F. Landis MacKellar  
& Wolfgang Lutz  
*Cambridge University Press: 2001. 266 pp.*  
£34.20, \$49.95

## **Norman Myers**

Despite marked declines in the annual growth of human numbers, the population explosion is far from over. The projection of a world population of 9 billion in 50 years will be 3 billion more than that of today, comparable to the increase of 3.5 billion during the past 50 years.

Much depends on how vigorously and urgently we tackle the problem of reducing human fertility. There are some 150 million married couples in developing countries who lack the contraceptive means to limit the size of their family. Were their birth control needs to be met — as they should be on humanitarian grounds even if there were no population problem — that ultimate global total of 9 billion would be cut by 1 billion people, at a cost of around \$3 billion per year. With the developing countries covering two-thirds of the cost — which they have agreed to do — the extra financial 'burden' per developed-world taxpayer would be the equivalent of a can of Coke every six months.

But for several years the rich countries have protested that they have never been poorer and cannot afford the cost, despite their earlier pledges in international forums. The concealed cost of the United States backtracking on its pledge at the 1994 International Conference on Population and Development, among other international forums, includes an estimated 4 million unwanted pregnancies per year. Fortunately, government treasuries, including that of the United States, have now reworked their arithmetic and have partially remedied the situation. But it largely remains a failed chance to turn a problem into an opportunity.

*Beyond Six Billion* presents the findings of a panel of population experts under the auspices of the US National Research Council. It provides a wealth of data and analyses, with emphasis on population projections, their assumptions, accuracy and uncertainties, and how demographic research is affecting projection procedures. The panel considers that projections up to 2050 are generally plausible.



Counting our blessings? Projections estimate that there will be 3 billion more of us in 50 years' time.

Fortunately, the book underscores the fact that demography is not destiny. Governments and citizens can change their population future. Indeed, many have been doing so, often with remarkable success. In 1984, at the start of its reinvigorated family-planning campaign, Kenya's fertility rate was 8.0 children per couple and its annual population growth rate was 4.0% — among the highest figures ever recorded for any country. In 2000, Kenya's fertility had dropped to 4.7 children, and its population growth rate was 2.1% — among the most rapid declines ever recorded.

Thanks in part to similar initiatives in other countries, the United Nations' 1994 medium-variant population projection for 2050 was reduced in the 1996 projection by a whopping 466 million people. But note the "in part". Population projections are also being modified by disease. The 1992 UN projection for sub-Saharan Africa in 2025 was 1,326 million people, whereas the 1998 projection proposed only 1,053 million. The difference is due to the impact of AIDS, whose demographic effects in several of these countries are proportionally equivalent to those of the plague in mid-fourteenth century Europe. And in 2010, the number of AIDS sufferers in sub-Saharan Africa will be overtaken by those in Asia. Nor are population plunges confined to developing coun-

tries. In 1992, Russia's 2025 projection stood at 171 million people, but the 1998 projection slumped to 137 million. This decline is due primarily to a fall-off in the country's health services.

Even with its recent drop in fertility, a country such as Kenya still suffers a substantial hangover from past population growth. It features a 'youthful population profile', with 45% of its citizens under the age of 15, compared with 31% in developed countries. This means that many potential parents are already in place. If from today these 'parents in waiting' have only two children per couple, the population will still grow for the best part of 50 years and all but double.

The book emphasizes a further phenomenon: the 'greying' of populations. By 2050 the proportion of people aged 65 and over in industrialized countries is expected to almost double, and in developing countries to triple. This will supposedly upset the traditional balance between healthy and less active citizens, and between productive and non-productive citizens.

But is this the best way to view the problem? Shouldn't we stop putting people out to grass on the Friday after their sixty-fifth birthday, and instead view them as a vast reservoir of experience and expertise — a resource rather than a burden? And by consigning the problem to the pending tray as

## book reviews

largely intractable, we are setting up a much more pressing problem for the future.

A probably still more important factor is that the book repeatedly speaks of a 2050 world in which huge numbers of people will still be in developing countries. Is this realistic? By 2025, China may well be the world's biggest economy, at least in terms of purchasing power parity. Already, at least 900 million people in so-called developing countries qualify as 'new consumers' with semi-affluent lifestyles. The perceived divisions of the world are becoming ever more blurred.

But despite these caveats, overall the book is an authoritative assessment of what we can expect for population totals at global, regional and national levels over the next half-century. It concludes that, for the most part, this will be a period when rapid population growth will slow to a virtual halt. The discussion of analytical methodologies is especially helpful to the non-demographer. The book might, however, have done more to stress that these are only demographic projections, and take no account of other factors, such as environmental constraints. Ethiopia has 65 million people today, many of them struggling to survive. Is it not likely that the country's mortality rate could rise, rather than decline as implied by most projections? Is it reasonable to expect that, within two generations, Ethiopia's population will grow to its projected 169 million? Perhaps it will, but the book should have highlighted potentially constraining factors.

The second book, *Population and Climate Change*, does indeed do just that. It considers environmental factors in terms of climate change, whether caused naturally or anthropogenically, and shows how climate change can induce demographic change, and vice versa. The fourteenth and fifteenth centuries featured unusually heavy rainfall and a high incidence of disease in Europe. Crop blights led to increased sickness in humans, with the result that the average longevity in England dropped within a century from 48 years to 38 years. The Black Death may well have originated in China or Central Asia during an exceptionally wet phase around 1332.

Today we are no longer so susceptible to most forms of established or conventional climate change, thanks to technological advances. But suppose that global warming causes the Gulf Stream to fail, as has been proposed in certain scenarios. This would profoundly disrupt agriculture, among other activities, in much of Western Europe. And the possible increase in the number and intensity of hurricanes along Florida's coasts would come on top of a fivefold increase in the state's population since 1950. Some 80% of today's populace lives within 35 kilometres of the coasts.

These and many other links between climatology and demographics are dealt with in an illuminating way by the three authors,

who are all at the International Institute for Applied Systems Analysis in Vienna. A first-rate exercise in multi-disciplinary analysis, the book examines a host of associated issues such as fertility shifts, population ageing, agriculture, energy consumption fostering greenhouse-gas emissions, health patterns and environmental security, together with the many policy options they entail, especially institutional adaptations. It amounts to a timely exploration of the major challenges ahead. ■

*Norman Myers is at Green College, University of Oxford, and at Upper Meadow, Old Road, Oxford OX3 8SZ, UK.*

## A scientist to count on

### **The Hilbert Challenge: A Perspective on Twentieth-Century Mathematics**

by Jeremy J. Gray  
*Oxford University Press: 2000. 328 pp.  
£20, \$34.95*

**W. Timothy Gowers**

It will come as news to many that the year 2000 was World Mathematical Year. Why was this great event not more widely recognized? Partly, of course, because mathematics is a minority interest, but also because mathematicians themselves did not agree on a single official way to celebrate. Several conferences claimed to be the main occasion on which leading mathematicians would take stock of their subject and predict its future, and there is more than one book with a similar purpose.

A century ago, mathematical futurology was dominated, indeed almost invented, by one person. At the age of 38, the German mathematician David Hilbert gave a lecture at the 1900 International Congress of Mathe-

maticians in Paris. In it he posed 23 problems, not all original to him, "from the discussion of which", as he put it, "an advancement of science may be expected". His lecture turned out to be extraordinarily influential — today's millennial conferences have all alluded to it, and could hardly have done otherwise.

One of the many virtues of Jeremy Gray's book about the Hilbert problems is that it convincingly explains how one relatively young man, giving one lecture — which, we are told, did not make much of an impression on its immediate audience — could have had the effect he did. One reason is that Hilbert's problems were very well chosen. He intended them to be difficult, but not so difficult as to be unapproachable, and, with one or two exceptions, his judgements were correct. Another reason is that Hilbert had worked in several areas — indeed, he is often held to be the last truly universal mathematician — so he was able to choose problems that spanned most of mathematics. A third reason is simply that the idea of presenting the world with a list of carefully selected unsolved problems was a brilliant one. Had Hilbert taken a more obvious path, such as outlining important programmes for research, he would not have captured the imagination of twentieth-century mathematicians in anything like the same way.

An important point that Gray brings out very well is that many of Hilbert's problems were linked by underlying philosophical and metamathematical themes. In other words, they were chosen not just for their intrinsic interest, but also for what they might reveal about the nature of mathematical argument itself. For example, it is a well-known fact of three-dimensional geometry that if two tetrahedra have the same heights and the same triangle as their base, then they have the same volume. But the only known proofs of this fact use calculus, or at least a limiting argument of some kind. By contrast, the equivalent two-dimensional theorem concerning areas of triangles can be proved by

## New in paperback

### **Vestal Fire: An Environmental History, Told Through Fire, of Europe and Europe's Encounter with the World**

by Stephen J. Pyne  
*University of Washington Press, £24*

### **The Art of Genes: How Organisms Make Themselves**

by Enrico Coen  
*Oxford University Press, \$16.95*  
"In *The Art of Genes*, Coen tries to explain to readers with no special knowledge of biology or development what is happening in the world of genetics ... I would have loved this book at 16, and so should anyone — aged 16 to 60 — who

really wants to understand development." John Maynard Smith, *Nature* **398**, 302–303 (1999)

### **Galileo's Daughter**

by Dava Sobel  
*Penguin, \$14*

"This is a harrowing tale of two victims: Galileo Galilei, forced to deny the evidence of his telescopes, and his daughter Virginia, forced at the age of 13 to enter the living sepulchre of the cloistered convent. Dava Sobel's latest well-timed bestseller provides a woman's angle on history's greatest conflict between science and religion." Brenda Maddox, *Nature* **403**, 702–703 (2000)



cutting one triangle into a few pieces and rearranging them to form the other one. Hilbert's third problem asked whether such a proof could be devised for tetrahedra. But his real interest was in the metamathematical question of whether the use of calculus was necessary. This was the first of his problems to be solved. In 1902, Max Dehn showed that calculus was needed.

Two of Hilbert's problems have, famously, had metamathematical solutions. His first problem was to prove or disprove Cantor's continuum hypothesis, which is the statement that there is no infinite set larger than the set of positive integers but smaller than the set of real numbers. Thanks to Kurt Gödel (in 1938) and Paul Cohen (in 1963), it is now known that this statement can be neither proved nor disproved. Hilbert's tenth problem asks for a systematic method for deciding which Diophantine equations have solutions. (Diophantine equations are polynomial equations whose solutions are required to be integers.) Building on the work of many mathematicians, Yuri Matiyasevich proved in 1970 that there was no such method. Results such as these have had a profound effect on the philosophy of mathematics.

The author of any mathematical book aimed at the general reader has to decide what background knowledge to assume, and Gray, like many others, is not consistent in his demands. This can be seen from a quick inspection of his 'boxes', those receptacles much loved of popular science publishers, which contain illustrations and (necessarily inadequate) explanations of some of the technical points in the text. That said, it would be misleading to describe this book as popular science. Two indications of its serious intent are that its title does not make silly use of the words 'history' or 'biography', and that we learn next to nothing about Hilbert's personal life. (For example, I still do not know whether he ever married.) As for the intended readership, at least some exposure to university-level mathematics is essential to appreciate the book properly.

My one complaint (apart from a few minor quibbles) is that Gray's prose contains far too many clumsily constructed sentences that I had to read twice. Here is one example from a long list: apparently, Hermann Minkowski thought it "unlikely that any polynomial in several variables which was never negative was expressible as a sum of squares". Surely, several of them are, one wonders, before realizing that the word "any" is supposed to be understood, unnaturally, as "every". This sort of writing lessens the pleasure of reading the book, which nevertheless remains illuminating and highly recommended. ■

W. Timothy Gowers is in the Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK.

## Going one better than nature?

**Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food**

by Vaclav Smil

MIT Press: 2001. 339 pp. \$34.95, £23.95

**John Emsley**

The greatest catastrophe that the human race could face this century is not global warming but a global conversion to 'organic' farming—an estimated 2 billion people would perish. That is the underlying message of this remarkable book, which charts the discovery of nitrogen fixation—the conversion of unusable atmospheric nitrogen to useful ammonia—and its impact on the world's food supply.

If crops are rotated and the soil is fertilized with compost, animal manure and sewage, thereby returning as much fixed nitrogen as possible to the soil, it is just possible for a hectare of land to feed 10 people—provided they accept a mainly vegetarian diet. Although such farming is almost sustainable, it falls far short of the productivity of land that is fertilized with 'artificial' nitrogen; this can easily support 40 people, and on a varied diet. Of course, 'organic' farming should be encouraged in order to recycle compost and dung. But it can never compete with the bountiful supply of agrochemical nitrogen, which now meets about 40% of the world's dietary needs.

Nitrogen is abundant in the atmosphere, but in a form that is difficult to extract; only a few microbes and plants have the capacity to do this. Yet, thanks to their efforts over aeons of time, a whole planetary ecology can now be sustained. This organic nitrogen will even



**Fritz Haber: discovered a way of converting nitrogen in the atmosphere into ammonia ...**

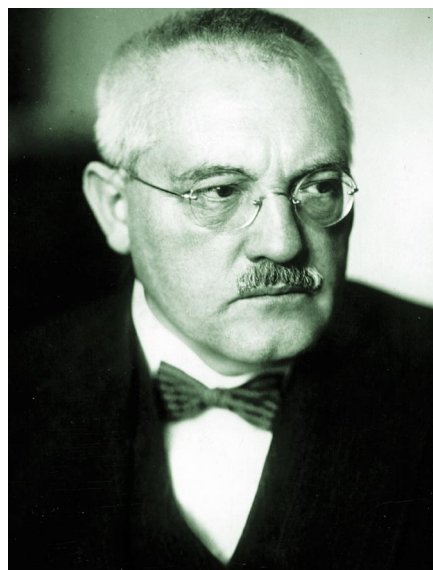
support continued agriculture if properly managed, but it imposes a maximum on the density of the human population.

All this changed on 3 July 1909, when two German chemists, Fritz Haber and Carl Bosch, proved that it was possible to convert atmospheric nitrogen into ammonia on an industrial scale. Today there are Haber-Bosch chemical plants around the world, producing 150 million tonnes of ammonia a year, most of which goes into making fertilizer. The nitrogen input into farmed land from these fertilizers now exceeds the natural input. Even low-income countries can afford Haber-Bosch factories, and these should begin to turn around food production there, just as they did in high-income economies.

In the final chapter of *Enriching the Earth*, Vaclav Smil of the University of Manitoba admits that he originally intended to write a biography of Haber and Bosch, but he quickly realized that an account of the effects of their research would be far more interesting, and concentrated on this. He was right to do so.

Smil begins by looking at the fact that all living things need nitrogen in order to make amino acids, the building-blocks for the proteins on which life depends. He explains how nitrogen is fixed naturally, and how traditional farming takes this from the soil, but with only partial success at returning waste material to fertilize future crops. The first successful nitrogen fertilizers came from the Chilean guano deposits in the nineteenth century, a clearly limited supply.

The central theme of *Enriching the Earth* tells of Haber's struggle to make hydrogen gas ( $H_2$ ) react directly with nitrogen gas ( $N_2$ ) to form ammonia ( $NH_3$ ), and of Bosch's faith that the process could be made to work commercially. Bosch then convinced the German chemical company BASF to invest in it. Thus was an industry born. But it was not immediately seen as the answer to the world's food



**... And Carl Bosch: had faith that the process could be made to work commercially.**

supply; instead, it fed into Germany's need for ammunition to fight two world wars. Ammonia from the Haber-Bosch factories was converted to nitric acid and thence to explosives. After 1945, however, the overwhelming use of such factories was to fix nitrogen for fertilizers.

Smil recounts how the industry developed, and how much of the world's population is now supported by it. He discusses how this chemical bounty is disbursed. Relatively little is used by US agriculture, but a great deal by Chinese farmers. Smil considers what will happen when developing economies also want their protein to be in the easily digested and tasty kind that comes from meat, even though this is the least efficient way of producing food. But can our planet support another 5 billion people on a Western diet, and won't more food simply encourage more humans to have yet more children? Smil's answer is found in his chapter "Nitrogen and civilization". The future looks surprisingly reassuring. The annual increase in global population will continue to decline even though food production is rising, and the total might well peak at less than 9 billion by the year 2050, declining thereafter.

This is a wonderful book, highly readable and replete with referenced data. It is soundly based on the chemistry that underpins our food supply, or at least the protein part of it, and is an ideal corrective to the misleading ideas we are constantly being fed by the organic food movement. Humans have a stark choice to make: do we farm four hectares of land 'organically' to feed 40 souls, or do we farm one hectare 'artificially', thereby leaving the other three to natural woodland and wildlife? There is a place for 'organic' farming, but only insofar as it permits us to recycle nitrogen that would otherwise go to waste. ■

John Emsley is in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

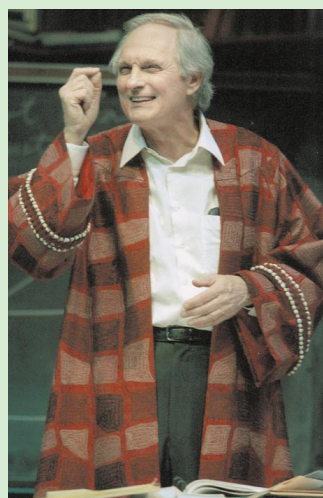
## Science in culture

### Enter Feynman, as clown

*QED*, a play by Peter Parnell, directed by Gordon Davidson. Horace Freeland Judson

The public Richard Feynman: O-rings, safe-cracking, bongo drums, naked women, atheism, the joyful questioning of authority, whether in physics or bureaucracies. The one thing everybody knows about him is that, when testifying before a panel of the US Senate investigating the Challenger disaster, he dropped a piece of O-ring, a rubberoid gasket, into a glass of ice water and demonstrated that such rings get brittle when cold — thus illustrating the cause of the fuel leak and explosion. Many have also read that, arriving at Los Alamos National Laboratory in 1943, a brilliant kid, PhD at 24, he could crack any of the offices' safes containing the secrets of the bomb research, thereby demonstrating that security was lax. And he played bongo drums, and was an obdurate, thoughtful atheist, and frequented topless bars in Los Angeles, and took art lessons so that he could draw nude models. Some also know that he was an inspired teacher at Caltech. The persona grew from interviews, public appearances and several books about him in which he connived. He played himself as trickster, an eccentric genius.

Enter Alan Alda as Feynman, rushing in to the bongo beat. Except for some voices on his answering-machine, and a brief appearance of a young woman student in Act II, this is a one-man show, and Alda plays Feynman broadly, as clown. The play is set in Feynman's office during an afternoon and evening in 1988, while he was dying, as we learn, of a crushing abdominal carcinoma. The office is cluttered, bongo up



front, desk piled, behind it a blackboard scrawled with mystifying mathematical jottings. Alda drums, jumps about the stage, drums, talks to the audience, to the telephone — his surgeon, his oncologist, this colleague or that — drums, piles up the Feynman anecdotes.

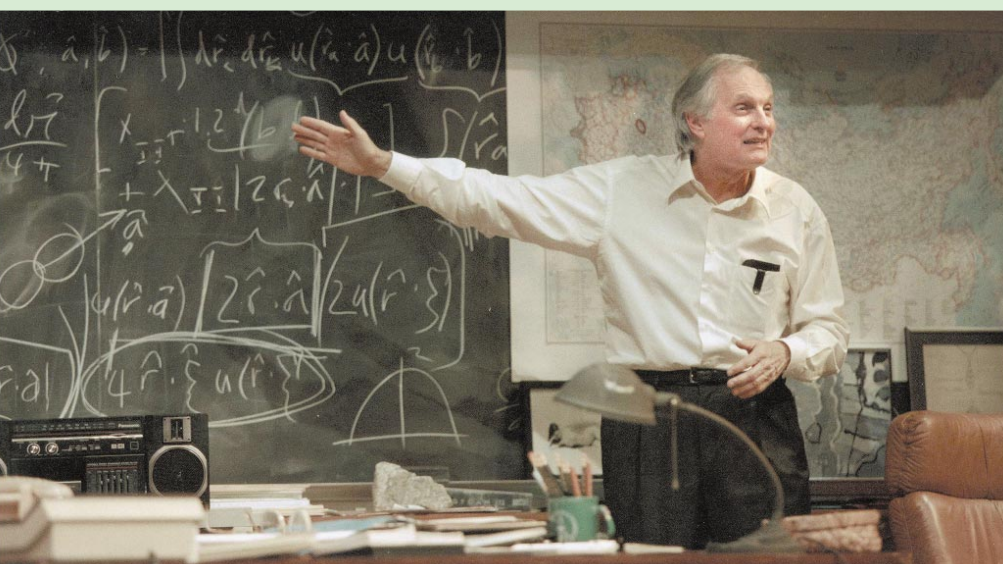
Alda is an actor celebrated for a variety of theatre and film work, but above all for 11 years in the TV series *M\*A\*S\*H*. His New York accent is perfect, he even looks acceptably like the man, his comic timing is

professional — but there are occasions in theatre when 'professional' is a substitute for depth, a term of reproach. What was behind the persona? Almost any photograph of Feynman shows his alert intelligence, twinkling with speculation, waiting to pounce. In the bruising intellectual world of high-energy physics, where everyone who makes it is an alpha male, Feynman's wit and assertiveness sprang from the originality and speed of his understanding. He applied this to the public world as well. Scientists know that he invented a graphic method, astonishingly original — Feynman diagrams — for anatomizing interactions of subatomic particles so that their behaviour could be calculated. This led to his work developing quantum electrodynamics, QED, the theory of almost everything, which won him a Nobel prize.

Actor, playwright, director — in this piece of theatre they have conspired to keep us from Feynman's intelligence. We get 40 seconds of a Feynman diagram scrawled on that blackboard as a gee-whiz illustration, when these were tools for discovery. Instead of Feynman's speed of comprehension we get frenetic, fussy movement. Elements of potential pathos — in Act I reminiscing about his first wife, dead of tuberculosis in 1945, in Act II facing the imminence of his own death — are drained of emotion by the constant straining for laughs.

What's left? An anthology of anecdotes and aphorisms. "Nature is a woman." "We must think in probabilities." "All science is a constant attempt to describe nature" — standing alone, a silly statement. "If you ask Nature the right question, she will give you the right answer." And at the end an embarrassing brief homily to that young student, to the effect that if she works really, really hard and can find the courage not to know certainty, she too can be a physicist. ■

Horace Freeland Judson is at the Center for History of Recent Science, George Washington University, Washington DC 20052, USA. *QED* is playing at the Mark Taper Forum, Los Angeles, until 13 May 2001.





# Heavenly phenomena

How an astronomer's words were transformed into a citation classic.

Euan Nisbet

The life of a scientific work proceeds by stages. First is an age of mouth and rumour, then it grows to abstract or preprint. Next it blossoms under critical review, then reaches the adulthood of formal publication. For most of our offspring, this brief springtime in the current printed issue lasts a week or a month, and is then followed by a high summer of a few citations within the next year or so. Next comes a lingering old age, surviving only in our own c.v. submitted annually to the dean. Eventually, the seventh age of the paper is in landfill.

But for some works (never our own), the last three ages of scientific life are different. They find fame, sooner or later, and are carried on in general knowledge by the citations of many, and in 1,000 student term papers. Then, as time passes, great works enter the texts as anonymous knowledge, important threads in the very fabric of science. And for the greatest knowledge, the fates do not cut off the threads — it becomes immortal, part of the received wisdom shared by us all, though its chronicler is long forgotten.

One of the finest works of Greek science is *Phaenomena*. This work was written by Aratus the Cilician, probably in the years following 276 BC. It summarizes the astronomical knowledge of its age. It has recently been edited, with a scholarly and erudite commentary, by the classicist Douglas Kidd. For many centuries, *Phaenomena* was one of the best known texts of science. Cicero translated it into Latin, as did Germanicus, the adopted son of Tiberius Caesar.

Most of *Phaenomena* is observational natural science as we know it today, not greatly different in purpose from modern textbooks. Much of the work is a map of the sky. The axis of rotation and pole are specified (north was then not quite where it is now), the constellations are carefully listed, and we are taught how to use them to measure time. Finally, we are given a lesson on the weather: words about clouds and wind. This is recognizably modern science: we are told to look for ourselves and “do not just keep up the observation carelessly”. Apart from a few wondrous introductory lines it is a detailed observational record, testable and useful.

Through Aratus we inherit the names of the north (*arctoi*, or Great and Little Bears), of the tropics of Cancer and Capricorn, and of the constellations. Aratus transmitted to us the knowledge of the great circle of stars that made the wheel of milk (*gala*), which the

poet John Milton later popularized as “the galaxy, that milky way”.

Unlike the sages of modern newsprint, Aratus did not dabble in myth, for he was an astronomer, not an astrologer. He saw and described, and invited the reader to test his observation. To Aratus, the 12 signs of the Zodiac and the five visible planets were features in the starry globe, to be used to define space and to measure time but not for divining the future. Milton, who was a member of Galileo's academy and associated with the founders of the Royal Society, is a true inheritor of Aratus.

The normal destiny of great texts is that the topics they discuss become invisibly written into the fabric of science. Aratus had a wider skill. He was not only a great text-writer, but a great poet. *Phaenomena* is written in poetry. Aratus takes technical material and puts it into lovely verse — imagine Keats rewriting Newton. Perhaps modern biochemistry texts should try poetry too, to make them easier to remember. Most of us start our papers with a couple of introductory lines, what the Greeks called a proem. Nowadays, our purple prose is usually a claim for the importance of our topic. Being a poet, and writing about the heavens, Aratus too allowed himself the luxury of a few brief lines before getting down to his catalogue of the stars: “With Dios let us begin.” This Dios (or Zeus) did not represent the planet Jupiter, nor even the misbehaving Greek god, but a wider universal God.

As poetry, *Phaenomena* was so admired that Virgil took Aratus as the direct inspiration for his *Georgics*, nature poems that are among the most fragrant of all Latin writings, and which, together with the direct influence of Aratus, helped inspire modern meteorology. Another who studied Aratus was a fellow Cilician, Sa'oul the Tarsean. Sa'oul was a thinker and poet. His poem on love, written to guide friends who lived in a raunchy boomtown, stands at the pinnacle of Greek literature.

Sa'oul was asked to address the Council of Athens, the embodiment of democracy. The council took its name from its original seat,

Aratus takes technical material and puts it into verse — perhaps biochemistry texts should try poetry.



Celestial view: Aratus gave us the ‘circle of stars’.

the Areopagos (the hill of Mars). The members enjoyed listening to new ideas. They thought Sa'oul was a *spermologos* (this spunky, typically Athenian insult self-translates into English), but they heard him out and even asked him back. Milton's essay *Areopagitica*, which sets out the logic that led to the First Amendment, honours the tolerance they showed.

In the speech, Sa'oul fused Hebrew insight with Greek philosophy: the speech marks the birth of Judeo-Christian culture. But which Greeks did he take? Not Plato or Aristotle. Instead, he picked two authors we would today welcome into the maths/science community. One is the shadowy figure of Epimenides, who, it is thought, had said of the deity: “In him we live and move and have our being”. Some propositions may be true but unprovable: Epimenides, the Cretan, gave us the paradox “All Cretans are liars”. Epimenides, with Solon, had helped create Athenian democracy, and, according to legend, had insisted on a statue to the unknown God.

After Epimenides, Sa'oul chose to add a line from Aratus's introduction to *Phaenomena*, a line also quoted by the Stoic Cleanthes. Roughly translated, it says “For we also are God's children”. And so, thanks to Sa'oul, *Phaenomena* contains probably the most printed citation in all scientific history. Perhaps it may dismay racists and creationists, but the message of support for observational, empirical science is clear. Sa'oul, also known as Paul, admired this textbook of natural science so much that, cited in the Christian bible, *Acts 17:28* — “For in him we live, and move, and have our being; as certain also of your own poets have said, ‘For we are also his offspring’” — are the words of the scientist Aratus.

Euan Nisbet is in the Department of Geology, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, UK.

## FURTHER READING

Aratus, *Phaenomena* (ed. Kidd, D.) (Cambridge Univ. Press, Cambridge, 1997).

# Why rename things?

Ralph A. Lewin

As an acquisitive small boy, when my pockets became too stuffed and the drawers of my desk could be opened only with difficulty, my father made me empty everything and sort it into categories: acid drops with aniseed balls, matchboxes with chrysalids, solder splashes with bits of lead, farthings with sous, and so on. In due course it comes naturally to humans to sort, categorize and name things. For ready retrieval, names can be alphabetized: fortunately, the order of letters is immutable.

Names can be descriptive, but where possible they too should be immutable for distinction and retrieval purposes. However stupid your neighbour, he's still *Homo sapiens*. A bald baby called Melanie (from the Greek word for black) may grow up as a blonde, but her name does not usually change on that account. John Smith may work in an office, and Bill Clark may shoe horses, but ordinarily Mr S. and Mr C. retain their names unchanged. That is why Linnaeus, in the middle of the eighteenth century, arranged that all animals and plants should have fixed binomials. What in previous centuries could have been *Rosa europaea spinosissima floribunda alba fragrans* (which would have described it) had its name fixed

succinctly as *Rosa alba* L. (for Linnaeus), even if it throws a pink variant. Even mistaken names — for instance, the North American beach vetch, *Lathyrus japonica* — are to be retained formally and technically.

Unfortunately, some microbiologists, who seem not to understand this neat idea, follow the medieval practice in describing, for instance, a little pink bacterium that turns paper mushy as something far too cumbersome, like *Microrhodobacillibacter cellulosolytica*. Entomologists, coping with a plethora of organisms among, say, beetles, tend to go in for neater names, such as *Ips*. (Microbiological neotaxonomists please note.) Incidentally, I think it's a good idea to spell out the names of genera on first use, as prescribed by the editors of *Nature*. For years I studied algae such as *Chlamydomonas elegans* without knowing that, among worm geneticists, *C. elegans* would refer to a completely different creature (although I don't think that either is particularly elegant).

Now that molecular biologists are producing phylogenetic trees (which, when congruent, we must consider irrefutable evidence for evolutionary pathways) a host of new problems arise. Because crocodiles are believed to be more closely related to birds than they are to turtles, for example, are sparrows just feathered reptiles, and does ornithology become merely a branch of herpetology? Hippos may be closer to whales than they are to their fellow ungulates such as pigs, so should librarians move hippo books down among the cetaceans? The lobe-finned coelacanth *Latimeria* is closer to humans than it is to herrings — so what price ichthyology? It is indeed a kettle of fish.

When it comes to higher orders, the plot thickens. Although we used to divide living things into plants and animals, some scientists have decreed that algae and fungi are not to be considered as plants. Their dictionaries told them that the Greek word for plant is *phyton*, so they renamed the Rhodophyta as Rhodophycophyta and the Cyanophyta as Cyanophycophyta. But stop a minute — if we synonymize prokaryotes with bacteria, then these blue-green microbes become Cyanobacteria. Moreover, since they can carry out photosynthesis, should we then call them photosynthetic bacteria? Traditionally, the latter don't evolve oxygen, and the blue-greens do. Hence they are neither phytoplankton nor phycoplankton, but maybe bacterioplankton. (Recently someone has thought up, and used, alas, the term Cyanoprokaryota.)

By similar arguments, the prokaryotic Actinomycetes, to be excluded from the eukaryotic fungi (*mycetes*), should be renamed Actinobacteria. And should all vas-

## Taxonomic classification

"Are sparrows just feathered reptiles, and does ornithology become merely a branch of herpetology?"

cular plants, having indubitably sprung from green algal ancestors, be now classified as highly differentiated members of the Chlorophyta?

Some scientists have decreed that fungi cannot have flagellated stages; so whereas rusts and smuts, like mushrooms, remain classed as fungi, the potato blight, for instance, has to go in with the algae. What criteria delimit algae, when molecular data now indicate that the brown algal kelps are closer to animals than they are to, say, the red what's-its-names? And if the Archaea are so clearly unrelated to, say, *Escherichia coli*, as they branch off so low in phylogenetic trees, should we stop calling them bacteria (as some claim), moving them from bacteriology into a separate discipline: archaeobacteriology or (even worse) archaeology?

Some of our colleagues, with little Latin and less Greek, have messed with the terminology of organs and organelles, too. Because wings were evolved by birds millions of years after they evolved among insects, and are unquestionably dissimilar in development and structure, then, not being homologous, they should be given a separate name — perhaps flapopodia, just as the flagella of algae, being wholly dissimilar in size and structure from the flagella of bacteria, have been renamed undulipodia. And then what should we call the Flagellata?

Actually, when it could be a matter of life and death, taxonomists in the medical profession do bypass molecular data that may indicate only a tiny proportion of genes distinguishing a pathogen from a harmless gut symbiont. Although they are probably indistinguishable under the microscope, the dangerous *Vibrio cholerae* is nominally distinguished from its benign cousins in the same genus.

Without denying the value of objective scientific data, unless they be demanded by expediency, I see no real benefit from adopting many such proposed taxonomic rearrangements or name changes. (Think back to Mr Clark and Mr Smith.) They muddle our language, and could introduce confusion in our libraries. Let's stick to widely accepted names and classifications whenever possible. ■

Ralph A. Lewin is in the Scripps Institution of Oceanography, University of California, San Diego, California 92093-0202, USA.

### FURTHER READING

Knapp, S. What's in a name? *Nature* 408, 33 (2000).



Widely accepted names enhance identification.

STEVE KAUFMAN/CORBIS

PATRICK JOHNS/CORBIS



# Climate and amphibian declines

J. Alan Pounds

Various reasons have been proposed for the falling numbers of amphibians in many parts of the world. Changing climate is likely to be a key factor — but with complicated links to the immediate causes of these population declines.

The shallow lakes and ponds of western North America provide plenty of sites where western toads (*Bufo boreas*) can lay their strings of eggs. But all is not well with these amphibians. In the crystal clear waters surrounded by snow-capped peaks in the Cascade range, the jet-black embryos are suffering devastating mortality. They develop normally for a few days but then turn white and die by the hundreds of thousands. On page 681 of this issue, Kiesecker, Blaustein and Belden<sup>1</sup> present evidence that climate change may be the underlying cause. Reductions in water depth due to altered precipitation patterns expose the embryos to damaging ultraviolet-B (UV-B) radiation, thereby opening the door to lethal infection by a fungus, *Saprolegnia ferax* (Fig. 1).

A sense of urgency surrounds the study of amphibian mortality. In the Cascades, western toads and various frog species have been decreasing in abundance since the mid-1980s. Similar declines, some leading to the extinction of entire species, have been reported from upland areas around the world<sup>2</sup>. Are the patterns a warning of environmental changes that may have increasingly unfavourable consequences for plant and animal communities, and for humankind?

Kiesecker *et al.*<sup>1</sup> show how important it is to consider various potential agents when



Figure 1 The end for embryonic western toads. Top, healthy eggs; below, the eggs infected with *Saprolegnia ferax*.

addressing this question. Previous studies in the Cascades, which examined the effects of UV-B exposure on embryos and its role in stimulating *Saprolegnia* outbreaks, pointed to stratospheric ozone depletion as a possible

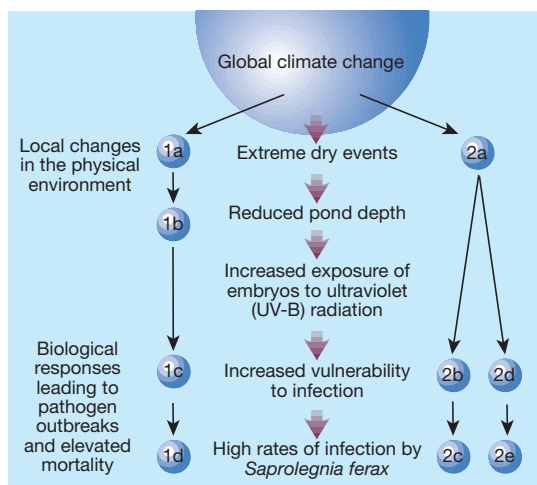
factor<sup>3,4</sup>. The new study, however, reveals that unusual weather, not ozone depletion, is the principal agent that increases exposure to UV-B. In extremely dry years, a large proportion of the western toad's egg-laying sites are in very shallow water, which provides little protection against UV-B. In water less than 20 cm deep, the white filaments of *Saprolegnia* invade and kill about 80% of the embryos, on average, compared to 12% in water deeper than 50 cm. Kiesecker *et al.* link the dry conditions to sea-surface warming in the Pacific, and so identify a complete chain of events whereby large-scale climate change causes wholesale mortality in an amphibian population.

Over the past 30 years, globally averaged air and sea-surface temperatures have risen sharply, and there is growing certainty that human activities are largely responsible<sup>5</sup>. A key region is the tropical Pacific Ocean, which has tended strongly towards warmer than average conditions since the mid-1970s<sup>6</sup>. Natural populations may be especially vulnerable where local climate is heavily influenced by the tropical Pacific. Together with long-term warming trends and changes in precipitation patterns, intense warm episodes of the El Niño/Southern Oscillation have conspired to produce extreme climatic events that have severely affected biological communities, from coral reefs to cloud forests<sup>7,8</sup>.

Western toads may likewise be victims of these events. Tropical Pacific climate influences winter precipitation (snowfall) in the Cascades, which largely determines water levels in lakes and ponds the following spring<sup>1</sup>. Precipitation totals are correlated with the Southern Oscillation Index for the preceding summer. The mass mortality of western toad embryos parallels the wholesale mortality of adult golden toads (*Bufo perigrinus*) and other amphibians in the mountains of Central America<sup>8</sup>. In both cases, alarming shifts in population patterns are linked to dry conditions associated with climate trends and fluctuations in the tropical Pacific.

Another recurring theme is epidemic disease. According to one hypothesis<sup>9</sup>, declines of highland amphibians are due solely to a skin disease caused by a chytrid fungus. This pathogen may indeed be a serious threat, but the single-disease model is likely to be an oversimplification. Some of the mysterious

Figure 2 Links between climate change and amphibian declines and extinctions. The central pathway outlines events leading to mass mortality of western toad embryos in the Cascade range of western North America<sup>1</sup>. The left- and right-hand pathways show other possible sequences. 1, Local increases in the day-to-day variability in precipitation (1a), together with atmospheric contamination, cause increased concentrations of toxic substances in microenvironments frequented by adult amphibians (1b)<sup>11</sup>. A resultant weakening of immune responses (1c) leads to high infection rates by a pathogen (1d) and high adult mortality. 2, Reductions in mist frequency in a tropical cloud forest (2a)<sup>8</sup> lead to high adult mortality through one of two biological sequences. Sequence 2b–2c is like that of 1c–1d, involving an increase in host susceptibility to infection. Alternatively, a pathogen's reproductive rates increase (2d) while host susceptibility remains unchanged, leading to increased rates of pathogen transmission (2e).



declines of frog populations in Central America and Australia have been accompanied by equally mysterious lizard declines<sup>8,10</sup>. A fungus that attacks moist-skinned amphibians is unlikely to attack reptiles, so the common denominator is probably broader than a single disease. Moreover, as Kiesecker *et al.* show, chytrids are not the only fungi killing amphibians in unprecedented numbers.

An alternative to the single-disease model is the hypothesis that extreme climatic events are causing amphibian declines by encouraging outbreaks of certain pathogens<sup>11</sup>. This idea is in line with the findings of Kiesecker *et al.* and with the growing recognition of the connections between climate and epidemics<sup>12</sup>. Pathogens may respond directly to weather, to physical factors influenced by weather, or to biological intermediaries, including changes in host susceptibility or in the vectors that transmit pathogens from one host to another.

Climate only loads the dice for disease outbreaks; it does not dictate when and where they will occur, and whether or not they will spread. It exerts its influence within the constraints of history, geography, natural history and potential interacting agents. Many amphibian populations are likely to suffer increased mortality in a warming world. There is no reason, however, to expect a close correspondence between climatic conditions and the occurrence of population declines.

Nor is there reason to expect a single chain of events whereby climate change leads to population declines (Fig. 2). The sequence of cause and effect that has resulted in the mass mortality of western toad embryos may apply to other shallow-water breeders with aquatic embryos. It might also apply where these species have suffered parasite-induced deformities<sup>13</sup>. But not all links in this particular chain apply to amphibian losses in general. In Costa Rica, Tilarán rain frogs (*Eleutherodactylus angelicus*) lay eggs in subterranean nests shielded from UV-B, but their numbers have nevertheless decreased. Many population declines, including those ascribed to chytrid outbreaks, have involved high mortality of adults rather than embryos. The association of chytrids with deaths of adult western toads in Colorado<sup>14</sup> illustrates a key point: a species or population might come under assault through several pathways.

Future studies of these sequences of cause and effect will help gauge the extent to which global climate change is driving amphibian population declines. During the 1990s, there was controversy over whether the declines were real, or simply a consequence of natural population fluctuations and direct human disturbances such as habitat destruction. Meanwhile, a separate debate focused on climate change and its relationship to green-

house-gas emissions. Today, there is little doubt that both phenomena — amphibian declines and global warming — are real. If there is indeed a link between the two, as the work of Kiesecker *et al.* suggests, there is clearly a need for a rapid transition to cleaner energy sources if we are to avoid staggering losses of biodiversity. ■

J. Alan Pounds is at the Golden Toad Laboratory for Conservation, Monteverde Cloud Forest Preserve and Tropical Science Center, Santa Elena, Puntarenas 5655, Box 73, Costa Rica.  
email: goldtoad@sol.racsa.co.cr

1. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. *Nature* **410**, 681–684 (2001).
2. Alford, R. A. & Richards, S. J. *Annu. Rev. Ecol. Syst.* **30**, 133–165 (1999).

3. Blaustein, A. R. *et al.* *Proc. Natl Acad. Sci. USA* **91**, 1791–1795 (1994).
4. Kiesecker, J. M. & Blaustein, A. R. *Proc. Natl Acad. Sci. USA* **92**, 11049–11052 (1995).
5. Intergovernmental Panel on Climate Change *Climate Change 2001: The Scientific Basis. Third Assessment Report, Summary for Policymakers. A Report of Working Group I.* (IPCC, Geneva, 2001).
6. Diaz, H. F. & Graham, N. E. *Nature* **383**, 152–155 (1996).
7. Reaser, J. K., Pomeroy, R. & Thomas, P. O. *Conserv. Biol.* **14**, 1500–1511 (2000).
8. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. *Nature* **398**, 611–615 (1999).
9. Morell, V. *Science* **284**, 728–731 (1999).
10. Czechura, G. *Wildl. Australia* **28**, 20–22 (1991).
11. Pounds, J. A. & Crump, M. L. *Conserv. Biol.* **8**, 72–85 (1994).
12. Epstein, P. R. *Science* **285**, 347–348 (1999).
13. Johnson, P. T. J., Lunde, K. B., Ritchie, E. G. & Launer, A. E. *Science* **284**, 802–804 (1999).
14. US Geological Survey News Release 9/13/1999.

### Cardiovascular biology

## Hearts and bones

Mark Sussman

The idea of repairing damaged heart tissue with donated cells is an old one, but finding cells that can do the job has been frustrating. The solution may come from a select group of bone marrow cells.

Heart attacks kill cardiac cells. In the aftermath, the damaged area of the heart forms scar tissue, which impairs cardiac performance. For millions of people every year, this progression from heart attack to heart failure serves as a grim reminder that current drug treatments are no substitute for the loss of living, beating heart cells. Ultimately, doctors resort to heart transplantation as a treatment for heart failure. Replacing the dead heart cells with living ones would be an attractive alternative, but has been stymied by biological, technical and ethical issues. Many of these problems are now circumvented by the demonstration by Orlic and colleagues<sup>1</sup> (page 701 of this issue) that a specific population of cells derived from bone marrow might hold the key to producing functional cardiac cells in damaged hearts.

Damaged hearts in animal models have been loaded with a staggering array of donor cell populations that are thought to be potentially useful for repair or remodelling<sup>2,3</sup>. Examples of donor populations include cells that form adult skeletal muscle, immortalized cells from the atrial region of the heart, smooth muscle cells, bone marrow cells and muscle cells of the heart (called cardiomyocytes), from embryonic, fetal and adult stages of development. Successful transfer depends on the donor cells surviving, maturing, electromechanically coupling with existing heart cells, and having a beneficial effect on the function of the recipient heart. This is asking a lot of any cell population, so it is not surprising that experiments so far have had varying results, depending in large

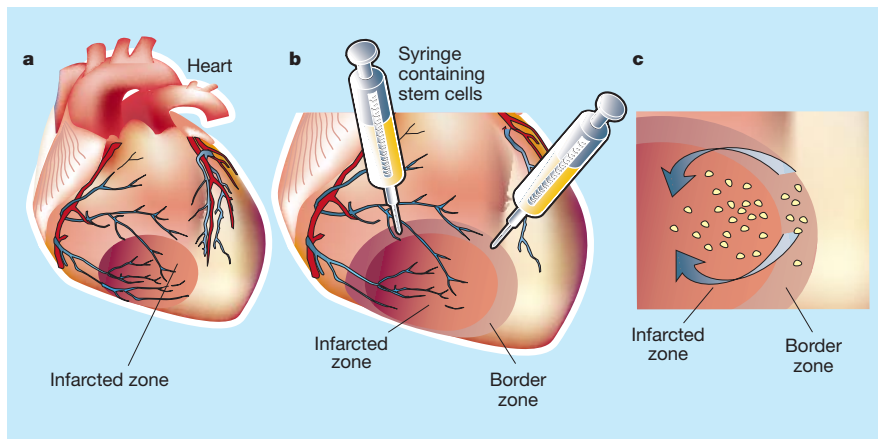
part upon exactly which donor cells are selected.

The idea of using bone marrow cells to repair the heart holds special allure. First, some marrow cells are multipotent — they are able to differentiate into several distinct cell types. In the context of the heart, such multipotent cells might be able to form heart muscle, as well as blood vessels to nourish the damaged area and promote its repopulation by muscle cells. Second, marrow cells are easily and routinely collected from adults, and so do not pose the ethical concerns inherent in using embryonic or fetal tissue. Third, treating a patient with cells derived from their own marrow eliminates the worry that the tissue will be rejected (a concern when using cells from another person). Fourth, transferring marrow cells into the scar tissue of a damaged heart is known to improve heart function, if the cells are first cultured for a week and then treated to induce the expression of muscle proteins<sup>4</sup>.

This combination of characteristics makes bone-marrow-derived cells uniquely suited to the task of restoring structure and function in the wake of a heart attack. The problem is that the main function of bone marrow is to generate blood cells. Multipotent cells in adults represent only a fraction of bone marrow cells, most of which are at various maturational stages on the way to forming the blood. Cells already committed to becoming blood cells will not become part of a reconstituted heart-muscle wall. Ideally, the multipotent cells would first be sorted out from the mixture, and then transferred to their new home in the heart.

Orlic *et al.*<sup>1</sup> describe the consequences of





**Figure 1** Repairing damaged heart tissue with bone marrow cells, as shown by Orlic *et al.*<sup>1</sup>. a, Ligation of the coronary arteries in a mouse heart causes loss of blood flow to a region of the heart, resulting in 'infarction' — damage to that part of the heart. b, Multipotent cells derived from the bone marrow of healthy mice are injected into the undamaged region (the border zone) surrounding the infarct. c, The bone marrow cells infiltrate into the damaged region, and start to express characteristics of heart muscle cells (cardiomyocytes), the smooth muscle that forms blood vessels, or the endothelial cells that line the vessel walls. Cardiac function improves.

doing just that in mice. They identified multipotent cells by the fact that they expressed on their surface a protein called c-Kit, but did not express the Lin protein. This combination of markers defines the multipotent population. (Very early, truly unspecialized multipotent cells are known as stem cells; whether or not the cells identified by Orlic *et al.* are stem cells remains controversial, but they are certainly multipotent.)

The authors isolated their donor cells from mice that had been engineered to express green fluorescent protein, so the cells could be readily identified when introduced into recipient mice that did not express that protein. Hearts of the recipient mice were damaged by ligation of the coronary arteries, causing loss of blood flow to a region of the heart-muscle wall, and localized tissue damage. After a few hours, the authors transferred donor multipotent cells into healthy tissue surrounding the damaged area (Fig. 1).

Within nine days, the transferred cells took up residence in the heart and showed remarkable adaptations to their new environment. Over half of the damaged area was filled with cells showing characteristics of cardiomyocytes, the smooth muscle that makes blood vessels or the endothelial cells that line the vessels. In other words, the transferred cells appeared to be differentiating to form the three major cell types needed to repair the heart. Orlic and co-workers also detected proteins associated with the activation of muscle-specific gene expression, indicating that some of the transferred cells were indeed on the path to becoming functional cardiomyocytes. New cells in the damaged region continued to multiply, and cardiac function improved. The implications are profound: damage to the heart muscle after a heart attack might be

reparable with specific bone marrow cells.

Predictably, these exciting observations raise a plethora of questions. The transferred cells may have expressed proteins that indicate they were adapting to the heart environment, but would the cells function properly after they had differentiated fully? Are the donated cells capable of establishing the architecture needed to form a functional myocardial wall? Can they repopulate scar tissue that forms within weeks or months after a heart attack, rather than only a few hours after loss of blood to a particular region? How long can the donated cells survive and continue to proliferate? And do these cells contribute directly to the recovery of the recipient heart, or do they merely stimulate the function of the undamaged heart tissue?

The search for cardiac stem cells has long occupied cardiovascular researchers hoping to find a way to replace damaged heart tissue. Such cells may not exist within the heart, but perhaps instead have been hiding within bones, masquerading as members of the blood-producing lineage. The challenge now will be to find out just how well these multipotent cells can perform new tricks in their new homes. Perhaps one day people suffering from heart failure might be treated so effectively that, as songwriter Paul Simon envisaged, "Their hearts and their bones can't be undone."

Mark Sussman is in the Division of Molecular Cardiovascular Biology, The Children's Hospital and Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA.  
e-mail: [sussman@heart.chmcc.org](mailto:sussman@heart.chmcc.org)

1. Orlic, D. *et al.* *Nature* **401**, 701–705 (2001).
2. Reinlib, L. & Field, F. *Circulation* **101**, E182–E187 (2000).
3. Kessler, P. D. & Byrne, B. J. *Annu. Rev. Physiol.* **61**, 219–242 (1999).
4. Tomita, S. *et al.* *Circulation* **100** (Suppl. II), 247–256 (1999).



#### 100 YEARS AGO

MR. H.G. Wells commences, in the current number of the *Fortnightly Review*, a series of speculative papers upon some changes of civilised life and conditions of living likely to occur in the new century. To construct a prehistoric animal from one or two fossil bones is a much easier task than the prediction of future developments from the point of view of the present; but Mr. Wells attempts to do this ... The subject of the first article is land locomotion in the twentieth century, and it scarcely requires a prophetic afflatus to know that the present systems will be largely superseded or modified. Horse traffic, with its cruelty and filth, while the animals exhaust and pollute the air, must give place to motor carriages in a few years. The railways will then develop in order to save themselves. There will be continuous trains, working perhaps upon a plan like that of the moving platform of the Paris Exhibition, or utilising the principle of the rotating platform outlined by Prof. Perry in these columns ... Nothing is said about the possibilities of aeronautics, not because of any doubt as to its final practicability, but because "I do not think it at all probable that aeronautics will ever come into play as a serious modification of transport and communication."

From *Nature* 4 April 1901.

#### 50 YEARS AGO

*Annual Review of Plant Physiology*. It is, perhaps, inevitable that, in a volume of 364 pages, a few errors should have escaped the diligence of the proof readers. Apart from trivial errors of spelling which will scarcely overtax the ingenuity of the informed reader, the generic name *Hibiscus* is spelt in a way which might surprise Virgil and Linnaeus (p.126). There is no real ambiguity in the use of the symbol  $O^{-2}$  for the oxygen anion (p.325); but it might have been helpful to distinguish on p.55 between published and unpublished data. It would be the height of folly, however, to attempt to judge the quality of fruit merely by inspection of the surface bloom. It would indeed be difficult to over-estimate the very great debt which we owe to the promoters of this new "Review". There is little doubt that the new series here inaugurated will prove indispensable to all those interested in the growth of plants.

From *Nature* 7 April 1951.

## Materials science

## Rapid alloy assessment

Robert W. Cahn

Combinatorial approaches provide a way of swiftly analysing huge numbers of samples. They are now being used to test the structural properties of materials, such as hardness and elasticity.

Combinatorial synthesis and screening is a way of automatically synthesizing 'libraries' of hundreds or even thousands of tiny samples of gradually changing composition, each to be tested for some property or other. The method, also known as high-throughput screening, was introduced by the pharmaceutical industry about a decade ago for the rapid assessment of potential drugs.

Since the mid-1990s, the same approach has been applied to the search for new materials with desirable properties, be they catalytic, luminescent, superconducting, magnetoresistant or ferroelectric. A 1995 paper by Xiang *et al.*<sup>1</sup> focusing on high-temperature superconductors stimulated rapid innovation, especially in industry, and there has been much subsequent work, both published and apparently in secret (see refs 2 and 3 for surveys). Xiang<sup>4</sup> has since produced a definitive overview of what he calls "combinatorial materials synthesis and screening of functional materials".

Writing in *Advanced Engineering Materials*<sup>5</sup>, Zhao now describes how he has applied the principles of the combinatorial approach to the assessment of structural (load-bearing) materials, such as superalloys used in the hot parts of jet engines. As a bonus, the paper describes a rapid method of determining phase diagrams. These diagrams are a central concept in metallurgy — they show

the crystalline 'phases', with their various compositions, that make up an alloy of particular overall composition, equilibrated at a constant high temperature.

High-throughput screening of functional materials depends on the production of libraries of thin films of a material, which are laid down in tiny squares, typically by evaporation or sputtering. These thin films cannot be used to determine the mechanical properties of structural materials, however. Instead, Zhao has adapted and improved an approach previously used for other purposes, which involves placing flat, polished pieces of metal or alloy in close contact. The pieces are then joined by an established technique known as hot isostatic pressing, in which a high hydrostatic pressure is applied through an argon atmosphere while the sample is heated.

Zhao's first example involves nickel and aluminium: pieces of Ni and NiAl were joined by high hydrostatic pressure for 4 hours at 1,200 °C. Prolonged further treatment at the same temperature followed to bring about extensive interdiffusion of the two metals. The resulting diffusion 'couple' was sectioned and polished, and the composition ranges of the various phases formed were analysed by electron probe microanalysis. Zhao then measured the local hardness of the phases with a mechanical nanoprobe. This instrument operates by measuring the

penetration of a minute diamond 'indenter' into a sample while the device is still under load. It provides data of high sensitivity and spatial resolution, and can determine hardness and elastic modulus at the same time<sup>6</sup>; measurements can be undertaken at high temperature if desired. The composition profile, phase diagram and hardness profile of the Ni–Al alloys are shown in Fig. 1.

Zhao has also applied the same principle to ternary or quaternary alloys of, for instance, iron, molybdenum and nickel. Here, three or four blocks in the form of sectors of a cylinder are welded together under pressure; multiple constituents are found in the interdiffusion zone near the point where the sectors meet. In the paper<sup>5</sup>, Zhao plots both the hardness and elastic modulus of one of the ternary phases as a function of molybdenum content after equilibration at 1,100 °C. In this way, progressive strengthening associated with increasing concentration of a solute (or two solutes) can be economically assessed. As Zhao describes in a forthcoming article<sup>7</sup>, these experiments can be repeated for different equilibration temperatures and alloy systems.

The interdiffusion approach produces what Xiang<sup>4</sup> has called a continuous phase diagram, because composition varies continuously, not in steps as in a multisample library. Attempts to use such diagrams as an economical way of determining isothermal sections in the phase diagrams of binary or ternary alloys go back a long way. Isothermal sections — that is, the graphical depiction of regions with single, double or triple phases present in equilibrium at a fixed temperature — are normally very laborious to determine, requiring the examination of dozens of discrete alloys.

The first serious attempt to produce continuous phase diagrams was made 36 years

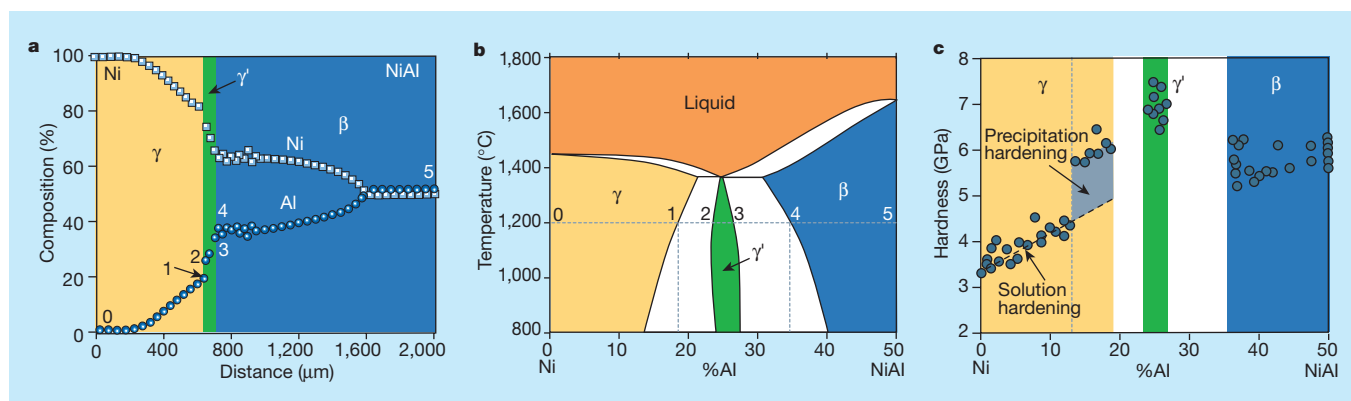


Figure 1 Composition profile, phase diagram and hardness profile of nickel–aluminium alloys (after ref. 5). These alloys are widely used in high-temperature applications. Pure Ni forms a solid solution with Al called the  $\gamma$  phase; likewise NiAl forms a solid solution with excess Ni known as the  $\beta$  phase. The intermediate phase ( $\gamma'$ ) is  $\text{Ni}_3\text{Al}$ . a, The composition profile of a diffusion couple, obtained by Zhao<sup>5</sup> with electron microprobe analysis, following treatment at 1,200 °C to bring about interdiffusion of the Ni and  $\text{Ni}_{0.5}\text{Al}_{0.5}$  samples. The numbers 1–5 cross-refer to part b of the figure.

b, The accepted Ni–Al phase diagram. The phases at 1,200 °C shown here match those determined by microprobe analysis, showing the accuracy of the microprobe technique. c, The results of hardness testing with the nanoprobe. In particular, at lower Al content in the  $\gamma$  phase the diagram reveals solid-solution hardening (caused by dissolving atoms of a different size in an element); at higher Al content in  $\gamma$ , there is additional precipitation hardening (due to fine precipitates of  $\gamma'$ ). White areas in these diagrams are mixed phases.



ago by Kennedy *et al.*<sup>8</sup>. They evaporated three constituent metals from the vertices of a triangle to form a continuously variable film on a substrate. But this approach did not give a simple spatial variation of composition. Much more recently, Xiang's group<sup>9</sup> developed a technique, based on sputtering through computer-controlled moving shutters, that produces an approximately linear variation in composition and so provides more useful continuous phase diagrams. The technique is not limited to metals. For instance, van Dover *et al.*<sup>10</sup> used a similar method, which they called 'the composition-spread approach', for high-throughput screening of thin-film insulating materials.

So far as determining phase diagrams is concerned, both the thin-film approach and the use of diffusion couples should be equally effective. But for assessing the structural properties of materials, only diffusion couples can be used. The nanoprobe can in principle also be used to determine resistance to creep — the slow deformation of a material under constant load and temperature. Several groups have shown that measurements of

the time dependence of the penetration depth of an indenter can provide reliable creep data.

This survey has shown that many rival names for combinatorial synthesis and screening, high-throughput screening and so on, have been put forward. The same happened a decade ago with the study of nanostructured materials, a field which is now flourishing. Perhaps only activities with high scientific promise undergo such intensive competition in nomenclature. ■

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.  
e-mail: rwc12@cam.ac.uk

1. Xiang, X.-D. *et al.* *Science* **268**, 1738–1740 (1995).
2. Service, R. F. *Science* **277**, 474–475 (1997).
3. Dagani, R. *Chem. Eng. News* 51–60 (8 March 1999); 66–68 (15 May 2000).
4. Xiang, X.-D. *Annu. Rev. Mater. Sci.* **29**, 149–171 (1999).
5. Zhao, J.-C. *Adv. Eng. Mater.* **3**, 143–147 (2001).
6. Oliver, W. C. & Pharr, G. M. *J. Mater. Res.* **7**, 1564–1583 (1992).
7. Zhao, J.-C. *J. Mater. Res.* (in the press).
8. Kennedy, K., Stefánsky, T., Davy, G., Zackay, C. F. & Parker, E. R. *J. Appl. Phys.* **36**, 3808 (1965).
9. Yoo, Y.-K. *et al.* *Nature* **406**, 704–708 (2000).
10. van Dover, R. B., Schneemeyer, L. F. & Fleming, R. M. *Nature* **392**, 162–164 (1998).

## Acoustics

# In a fly's ear

Peter M. Narins

Organisms often identify the source of a sound by comparing the noises that arrive at the two ears. Using some interesting tricks, a minute fly has mastered this feat as accurately as humans.

The auditory systems of animals from insects to mammals are adept at detecting and identifying sounds — and, crucially, at locating their source. Although it is certainly helpful for a female tree frog to be informed of a male's presence in the vegetation from the sound of its call, it is arguably more useful to be able to locate the source of the call with precision, so that (in this case) mating can occur. Similarly, when the small

parasitoid fly *Ormia ochracea* hears the song of the field cricket (*Gryllus* species), it must be able to locate the exact source of the song: female flies deposit larvae on or near the calling crickets, which ultimately serve as a food source for the developing fly larvae. Writing on page 686 of this issue<sup>1</sup>, Mason and colleagues reveal the degree of precision with which *Ormia* can locate crickets from their songs. They also describe a neuronal mecha-

nism that could underlie this remarkable behaviour.

In general, the position of a sound source in three-dimensional space is computed by the auditory system from two physical cues: the difference in the time it takes for the sound to reach each of the ears (interaural time difference), and the difference in the intensity of the sound that reaches the ears (interaural intensity difference). These cues vary according to where the sound source is located relative to the midline axis. The midline axis runs equidistant between your ears, from behind to in front of your head. The interaural time difference is zero for a sound source located on the midline, but increases with the angular displacement of the sound source from the midline. (For a sound source that is located directly to the right or left of your head, the angular displacement is 90°.)

Sound localization becomes especially challenging for small animals such as *Ormia*, for which the distance between the ears is only 0.5 mm. In practice, this means that the maximum interaural time difference for *Ormia* (that is, when the sound is located at 90° to the midline) is only about 1.5 µs, and the interaural intensity difference is vanishingly small. Even so, these flies can pinpoint sounds exceptionally well. Mason *et al.*<sup>1</sup> report that *Ormia* is capable of locating a loudspeaker broadcasting cricket songs to within 2° of the midline. This means that *Ormia* can detect a change in interaural time difference of a mere 50 ns. How does it achieve this?

The tympana (eardrums) of *Ormia*, located behind the fly's head, are connected internally by a cuticle-based bridge that functions as a flexible lever<sup>2,3</sup>. This unusual structure means that the membranes of the tympana can vibrate in response to sound in two distinct ways, with different resonant frequencies<sup>3</sup>. Over the operating range of 5–20 kHz, the actual motion of the tympana is the result of a linear combination of these two resonant modes. There are two consequences of this arrangement<sup>2,3</sup>. First, the

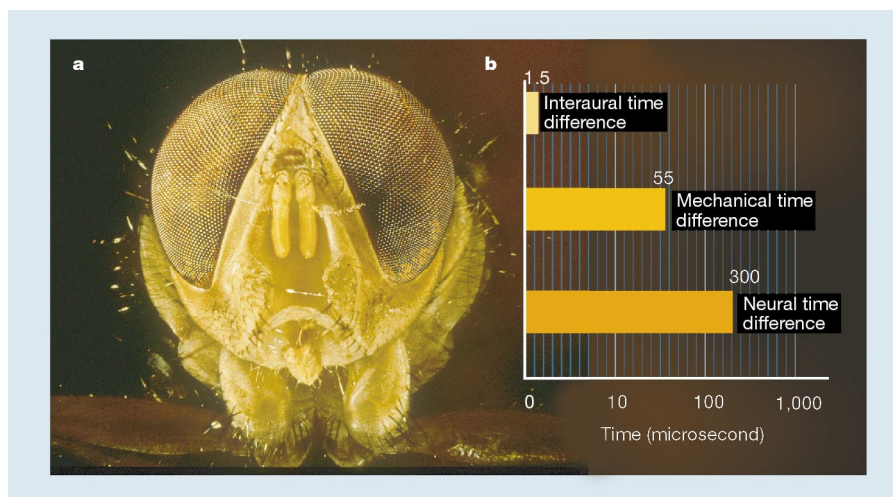


Figure 1 The parasitoid fly *Ormia* can locate cricket songs with pinpoint accuracy. It can do so by amplifying the difference between the sounds that reach each ear. a, The head of *Ormia*. b, Amplification of sound differences. The interaural time difference refers to the maximal difference between the amount of time it takes sounds to reach each ear. The mechanical time difference is the maximal difference between the times at which the two eardrum membranes respond to the sound. The neural time difference is the maximal difference between the response latencies of the left and right auditory nerves. At each stage the left–right differences are exaggerated, enabling the central nervous system of the fly to localize cricket songs with stunning accuracy, as shown by Mason *et al.*<sup>1</sup>.

maximum difference in the amplitude of vibration of the two eardrums increases to 10 dB. Second, the maximum time delay between the mechanical displacements of the two eardrums increases to about 55  $\mu$ s. So, minute interaural time differences are converted by *Ormia*'s unique tympanal structures into respectable interaural amplitude differences and increased interaural mechanical time differences. These differences can be processed more easily by sensory receptor neurons.

Mason *et al.*<sup>1</sup> have now measured the differences in the time it takes for individual sensory receptor neurons to fire (a time known as the latency) in response to sounds at various angles from the midline. They also measured the 'summed response latency differences' for the whole auditory nerve — the collection of sensory receptor neurons — in response to stimuli from the left and right sides of the animal. Neuronal response latencies are known to shorten in direct proportion to the intensity of a stimulus<sup>4</sup>, so the neuronal latency differences for sounds impinging on the fly from 90° to the midline are six times higher (about 300  $\mu$ s) than the mechanical time differences<sup>1</sup>. In other words, the interaural difference increases still further with this next step in sound detection (Fig. 1).

Nevertheless, Mason *et al.* find that interaural neuronal latency differences are only about 3.5  $\mu$ s per degree of angular displacement. So, when the source of the sound (the loudspeaker) is moved through an angle of 2°, the corresponding latency differences are remarkably small, about 7  $\mu$ s. Adding insult to injury, the variability in the timing of firing by receptor neurons following repetitive stimuli is ten times greater than this, of the order 70  $\mu$ s. Mason *et al.* suggest that, for *Ormia* to detect a latency (signal) of 7  $\mu$ s in jitter (noise) of 70  $\mu$ s, the fly must use a mechanism of temporal hyperacuity, much like that first described for the weakly electric fish *Eigenmannia virescens*<sup>5,6</sup>. Temporal hyperacuity is the result of convergence of many receptor neurons onto a central cell. Each input neuron, when it fires, generates a subthreshold action potential in the central cell. That cell will fire an action potential only when it receives most of its inputs together. This arrangement is relatively resistant to 'outlier' subthreshold potentials that occur much earlier or later than the mean. So the temporal variability in the firing of action potentials by the central cell will be reduced.

Despite their inherent noise, sensory systems often have specialized mechanisms for reliably detecting extremely small signals. One such mechanism, stochastic resonance, takes advantage of random, broad-spectrum noise in the sensory receptors to increase the system's sensitivity to weak signals. Not surprisingly, stochastic resonance has been

described in several sensory systems, including mechanoreceptors<sup>7–9</sup>, photoreceptors<sup>10</sup>, electroreceptors<sup>11</sup>, auditory hair cells<sup>12</sup> and auditory nerve fibres<sup>13</sup>. Work on *Eigenmannia*<sup>5,6</sup> and now *Ormia*<sup>1–3</sup> has addressed the problem of the temporal variability that is introduced by sensory receptors. It seems that temporal hyperacuity is one mechanism that may help to overcome such variability, and so might be seen as complementary to stochastic resonance. It is only because its central nervous system reduces temporal variability that *Ormia* is able to localize sound sources to within 2°. Ineluctably, further studies of this remarkable insect ear will lead to the development of biologically based, miniaturized directional 'ormiaphones' for potential use in hearing aids and other listening devices. ■

Peter M. Narins is in the Department of Physiological Science and the Brain Research Institute, University of California, Los Angeles,

California 90095-1606, USA.

e-mail: pnarins@ucla.edu

1. Mason, A. C., Oshinsky, M. L. & Hoy, R. R. *Nature* **410**, 686–690 (2001).
2. Miles, R. N., Robert, D. & Hoy, R. R. *J. Acoust. Soc. Am.* **98**, 3059–3070 (1995).
3. Robert, D., Miles, R. N. & Hoy, R. R. *J. Comp. Physiol.* **179**, 29–44 (1996).
4. Stevens, C. F. *Neurophysiology: A Primer* (Wiley, New York, 1966).
5. Rose, G. & Heiligenberg, W. *Nature* **318**, 178–180 (1985).
6. Kawasaki, M., Rose, G. & Heiligenberg, W. *Nature* **336**, 173–176 (1988).
7. Douglass, J. K., Wilkens, L., Pantazidou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
8. Collins, J. J., Imhoff, T. T. & Grigg, P. *J. Neurophysiol.* **76**, 642–645 (1996).
9. Levin, J. E. & Miller, J. P. *Nature* **380**, 165–168 (1996).
10. Barlow, R. B., Birge, R. R., Kaplan, E. & Tallent, R. *Nature* **366**, 64–66 (1993).
11. Braun, H. A., Wissing, H., Schäfer, K. & Hirsch, M. C. *Nature* **367**, 270–273 (1994).
12. Jaramillo, F. & Wiesenfeld, K. *Nature Neurosci.* **1**, 384–388 (1998).
13. Narins, P. M., Benedix, J. H. Jr & Moss, F. *Aud. Neurosci.* **3**, 389–400 (1997).

## Physical chemistry

# Molecules at the edge

Peter J. Rossky

Ultrafast observations of fluid interfaces show that the way molecules interact at these junctions is both structurally and dynamically different from interactions in the bulk liquid.

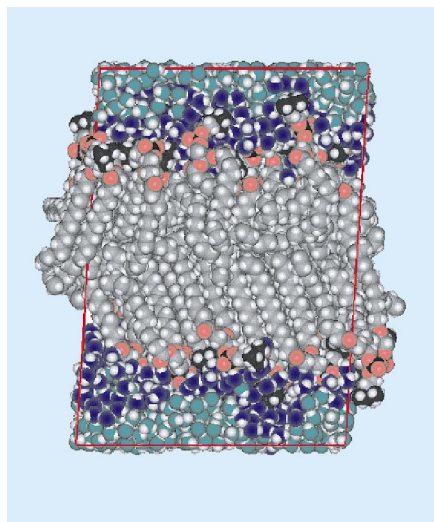
The bulk properties of solutions depend very much on the ability of dissimilar molecules to participate in attractive interactions without forming chemical bonds — a process known as solvation. For example, the solubility of sodium chloride (the solute) in water (the solvent) can be attributed to favourable interactions between the surrounding water and each sodium and chloride ion. But how does solvation change where two different fluids meet? Such interfaces are found in many biological and industrial chemical systems, and are crucial in creating useful compartments for different chemical processes and environments. In a report in the *Journal of Physical Chemistry B*, Benderskii and Eissenthal<sup>1</sup> provide us with a direct view of solvation at an important type of aqueous interface. By using femtosecond ( $10^{-15}$  second) spectroscopy, they are able to measure the effects of a membrane-like molecular layer on the solvation of a dye molecule embedded at the water interface.

Biological processes in which solvation plays a crucial role include electron transfer across membrane-bound proteins, and ion and proton transport through membranes. Related processes can occur in chemical environments, such as soap solutions and microemulsions. Solvation is important in all of these processes because the movement of charges needs to be counterbalanced by

the rearrangement of the solvent molecules. For charge transfer to proceed quickly, the energy cost must be small, so the solvent and the charge must follow one another closely. For the lightest charged particles (electrons and protons), the reorganization of the heavier solvent can be the rate-limiting step. To measure the speed at which the solvent responds to charge migration (solvation dynamics), observations that are sensitive to molecular movements — operating on femtosecond timescales — are needed.

Benderskii and Eissenthal<sup>1</sup> examine the dynamics of an aqueous interface by using a dye molecule as a probe solute. Bulk water is one of the most important solvents, and, owing to its small molecular size and the small masses of its atoms, also one of the fastest. To record the dynamical behaviour of liquid water the authors use an ultrashort laser pulse to excite the dye, changing its molecular charge distribution. As a result, the solvent, which is initially in equilibrium with the dye molecules, finds itself in a non-equilibrium configuration with respect to the newly excited state. The water molecules respond by moving to create a new equilibrium state of solvation. This evolution is monitored on a femtosecond timescale by another ultrashort laser pulse, using a technique that senses only those molecules at the interface (so-called second harmonic generation).





**Figure 1** Molecular model of a lipid bilayer–water interface. The lipid bilayer is dipalmitoylphosphatidylcholine. Bulk water molecules are green, whereas the molecules most involved in solvation — forming favourable interactions with neighbouring hydrophilic lipid atoms — are blue. The hydrophilic lipid atoms are highlighted in black ( $\text{N}(\text{CH}_3)_3$  groups and phosphate phosphorus) and red (phosphate oxygen and carbonyl oxygen). The structure of the interface is distinct from that of bulk water or an air–water interface, in that the hydrophilic groups are well mixed with the water molecules. In their experiment, Benderskii and Eissenthal<sup>1</sup> observe the behaviour of a solute at a similar interface between a lipid-like monolayer and water. (Courtesy of D. Tobias, Univ. California, Irvine, and M. L. Klein, Univ. Pennsylvania.)

Benderskii and Eissenthal are particularly interested in what happens at the interface when the water surface is covered with a lipid-like monolayer of protonated fatty acid. This membrane-like layer contains both hydrophobic and hydrophilic regions, pointing towards the air and water respectively. The effect of the lipid-like layer on the properties of the interface can be monitored in two ways. First, the strength of equilibrium solvation of the dye at the interface can be measured from the size of the frequency shift in the dye absorption spectrum. Second, the solvation dynamics can be tracked using ultrashort laser pulses. Benderskii and Eissenthal find that both the equilibrium solvation and the solvation dynamics are sensitive to the presence of the lipid-like layer.

Structurally, the hydrophilic (protonated acid) ends of the lipid-like layer are densely clustered at the interface with water, so one might expect this interface to be uniquely organized. Indeed, Benderskii and Eissenthal show that the equilibrium solvation of the dye molecules at the lipid interface is of intermediate strength between bulk water and an air–water interface. Figure 1 shows a ‘snapshot’ from a molecular simulation of a similar environment — a lipid bilayer interface with water<sup>2</sup>. In this membrane-like configuration, the interface with water has a mixed molecular character, with hydrophilic groups (black and red) interspersed with neighbouring water molecules (blue).

The lipid-like layer can also have an effect on solvation dynamics. At a bare air–water

interface, solvation dynamics are similar to those in bulk<sup>1</sup>. But Benderskii and Eissenthal show that solvation occurs roughly twice as fast at the interface with the lipid-like layer as it does in bulk water or at an air–water interface.

The response time of the solvent is not simply considerably faster with the lipid-like layer — it is also different. The diffusive response time of bulk water can be described by two components: a faster solvation component (of around 250 femtoseconds) and a slower component (around 1,200 femtoseconds). With the lipid-like layer, the faster component actually increases to about 400 femtoseconds, but the slower component is suppressed. The net result is a shorter overall solvation time. The faster component seen at the air–water interface or in bulk water is mostly associated with molecular reorientation, whereas the slower component is more associated with collective diffusion of the solvent. The authors explain the suppression of this slower component by the lipid-like layer as the result of interfacial interactions that break up the extended network of water hydrogen bonds close to the interface. This would reduce the tendency of the solvent to respond collectively, reducing the importance of the slower collective motion.

The hydrophilic groups in the lipid-like molecules can themselves act as a solvent, so it is best to think of the lipid–water interface as an alternative solvent, rather than as a perturbed water environment. The change in solvation dynamics, relative to bulk water or the air–water interface, is consistent with this point of view. The authors note that the sol-

## Materials science

### Turbulent creep

Applying long-term, constant stress to ice causes it to deform like a plastic, meaning that it won't spring back to its original shape. The deformation has typically been thought to be a smooth dynamical process similar to the laminar flow of fluids. But on page 667 of this issue, M.-C. Miguel and co-workers show that the way ice flows under stress is actually turbulent.

Miguel *et al.* use a technique called acoustic emission — recording of acoustic waves emitted during plastic deformation — to measure the internal dynamics of squeezed ice. They find that one-dimensional defects in the crystalline lattice, known as dislocations, move in intermittent bursts. And from the particular statistical distribution of the

energies associated with each acoustic burst (it follows a power law) they conclude that plastic deformation in ice is more akin to turbulent flow than laminar flow.

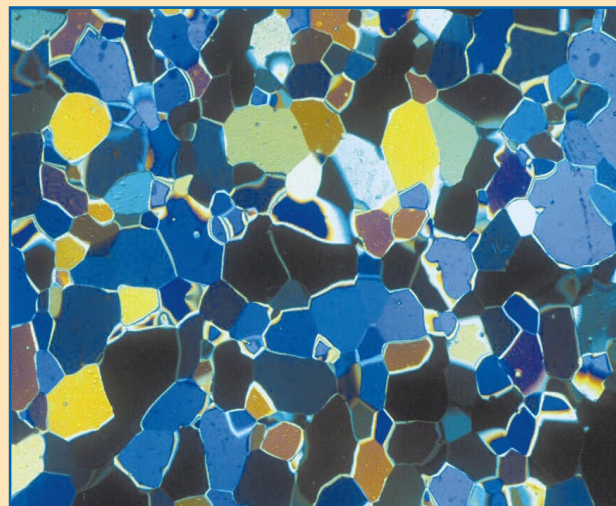
The textbook view of plastic deformation is that the dislocations glide and slip along smoothly. But this is true only of average behaviour over large length scales. A close look at artificial granular ice reveals a complex and unpredictable structure, as the image on the right illustrates. These crystals are enhanced by polarized light and the average grain size is about 1.2 micrometres.

To confirm that the effect they have identified is not specific to ice, Miguel and co-workers have done numerical simulations; what they have found is that turbulence should

be considered a general feature of creep flow. These results should have implications for our understanding of plastic

deformation more generally, and have practical bearing on problems in materials science and engineering.

Josette Chen



vent response time is directly related to the freedom of molecular motion in the environment. It is important to keep in mind that such freedom need not follow the strength of individual interactions. It is the balance of opposing forces — not their individual values — that determines this freedom, with perfectly balanced forces in all directions allowing the most freedom. Hence, observations of the strength and dynamics of solvation are largely complementary, and here indicate that the interface is a new environment.

It is worth noting that rates of charge transfer can often be expressed without invoking dynamics at all. Alternative approaches consider the probability of reaching a critical intermediate (transition) state between reactant and product charge distributions. This probabilistic description can be successfully applied to electron transfer<sup>3</sup>, as well as to descriptions of ion pairing and separation. The key to these approaches is that the solvent configuration is an explicit

factor in the path leading from reactants to products. The closely tied movements of charge and solvent remain a crucial element.

Whether the observed differences in the solvation behaviour between membrane-like interfaces and bulk water play an important role in biological functions at membranes is not yet clear. Many important factors, such as transmembrane electrochemical potentials, are governed by processes operating on much longer timescales and length scales than reflected in these measurements. Nonetheless, a successful description of the molecular mechanisms behind chemical processes at interfaces will almost certainly include molecular solvation as well as larger-scale factors. ■

Peter J. Rossky is at the Institute for Theoretical Chemistry, University of Texas at Austin, Austin, Texas 78712, USA.

e-mail: rossky@mail.utexas.edu

1. Benderskii, A. V. & Eisinger, K. B. *J. Phys. Chem. B* **104**, 11723–11728 (2000).
2. Tobias, D. J. *NATO Sci. Ser., Ser. A* **305**, 293–310 (1999).
3. Suttn, N. *Adv. Chem. Phys.* **106**, 7–33 (1999).

## Cell biology

# Channelling calcium

James W. Putney Jr

Rises in the level of calcium ions inside cells are key biological signals. A long-sought protein involved in the 'store-operated' process by which calcium ions enter cells has now been identified.

Calcium ions are important biological signals, controlling processes such as protein secretion, muscle contraction, cell death and development. Calcium signalling involves an increase in the concentration of calcium ions inside cells, and one of the mechanisms by which this occurs is called capacitative calcium entry, also known as store-operated calcium entry<sup>1</sup>. This process requires the regulated opening of ion channels in the cellular outer membrane, the plasma membrane. But it is not certain which molecules constitute the channels, and until these molecules are identified, our understanding of this important process will remain limited. On page 705 of this issue<sup>2</sup>, Yue and colleagues present evidence implicating a newly discovered protein, CaT1, as a constituent of the ion-conducting pore of a capacitative calcium-entry channel. Their discovery may lead to a deeper understanding of the cellular and molecular mechanism by which this channel is controlled.

Hormones and neurotransmitter molecules trigger increases in the concentration of  $\text{Ca}^{2+}$  ions inside cells by one of two basic mechanisms. They may cause the release of  $\text{Ca}^{2+}$  from stores within the cell, or they may activate ion channels in the plasma membrane, allowing  $\text{Ca}^{2+}$  to enter from the out-

side. Among the various mechanisms by which the regulated entry of  $\text{Ca}^{2+}$  occurs, one of the most common is capacitative or

store-operated  $\text{Ca}^{2+}$  entry<sup>1</sup>. This process is so called because it is somehow (Fig. 1) activated by a fall in the concentration of  $\text{Ca}^{2+}$  ions stored in the endoplasmic reticulum, one of the cellular organelles. Such a fall generally (but not always) occurs in response to a lipid-derived messenger molecule, inositol-1,4,5-trisphosphate ( $\text{IP}_3$ )<sup>3</sup>. Consistent with its nearly ubiquitous nature, store-operated  $\text{Ca}^{2+}$  entry seems to be essential in such diverse cellular responses as muscle contraction, protein secretion, metabolism, cell differentiation, cell division and programmed cell death. An understanding of the molecular and cellular mechanisms underlying capacitative  $\text{Ca}^{2+}$  entry would be of enormous benefit to basic and applied biomedical research in several areas.

Lately, researchers in this field have looked particularly at two interrelated issues: the identity of the capacitative  $\text{Ca}^{2+}$ -entry ion channels, and the nature of the signal that activates them. Solving the first problem would provide a good starting point for tackling the second. From electrophysiological experiments, researchers know how the channels behave, but their molecular components have so far remained a mystery.

The fruitfly *trp* gene and its mammalian counterparts have been considered likely to encode components of capacitative  $\text{Ca}^{2+}$ -entry channels<sup>4</sup>. But, although several papers have produced evidence for this view, several other reports provide conflicting data<sup>5</sup>. In most instances, the known TRP channel proteins are activated by mechanisms other than depletion of  $\text{Ca}^{2+}$  stores, and, perhaps more significantly, none of the

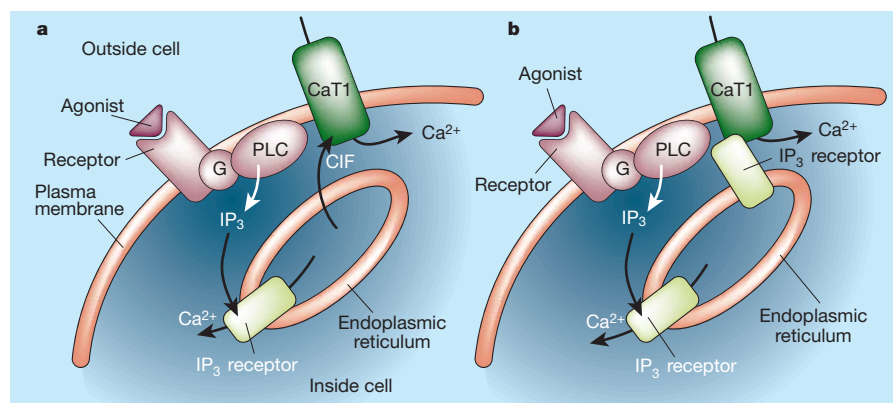


Figure 1 Two models for how the depletion of intracellular calcium stores might trigger the influx of calcium into a cell. In many cell types, activation of a cell-surface receptor by its agonist (that is, a neurotransmitter or a peptide hormone) results in the production of inositol-1,4,5-trisphosphate ( $\text{IP}_3$ ) (via a G protein and phospholipase C, PLC).  $\text{IP}_3$  binds to its receptor on the endoplasmic reticulum, resulting in depletion of stored  $\text{Ca}^{2+}$ . a, According to one theory<sup>12</sup>, following discharge of the endoplasmic-reticulum  $\text{Ca}^{2+}$  stores, a diffusible signal (termed calcium influx factor, CIF) is released from the endoplasmic reticulum, and binds to and activates store-operated  $\text{Ca}^{2+}$  channels on the plasma membrane. b, The conformational coupling model<sup>13,14</sup> proposes that discharge of  $\text{Ca}^{2+}$  stores leads to a conformational change in the  $\text{IP}_3$  receptor, which can then bind to plasma-membrane  $\text{Ca}^{2+}$  channels. Yue *et al.*<sup>2</sup> have identified the protein CaT1 as a likely component of a plasma-membrane, store-operated  $\text{Ca}^{2+}$  channel. This discovery may allow these and other models for store-operated  $\text{Ca}^{2+}$  entry to be tested.



*trp* genes has been shown to encode a channel with the precise ion-conduction properties expected of store-operated channels. Indeed, in most (but not all) cases, the channels produced by experimental expression of *trp* genes appear to be nonselective cation channels, whereas the capacitative  $\text{Ca}^{2+}$ -entry channels are thought to be among the most selective  $\text{Ca}^{2+}$  channels<sup>6</sup>.

There are seven close mammalian relatives of the fruitfly *trp* gene, and there is also a much larger family of more distantly related genes. This larger family can be divided into three subfamilies, based on similarities in the structures of the encoded proteins<sup>7</sup>. The fruitfly TRP protein and its seven close relatives are 'short' TRP channel proteins. Members of a second group are 'long' TRP channel proteins.

The third group comprises 'osm' channels, so called because they appear to be regulated by physical stimuli such as osmotic concentration, heat or mechanical stress. Two members of this family — called ECaC and CaT1 — were first identified as a result of searches for channels involved in transporting  $\text{Ca}^{2+}$  across epithelial cells<sup>8,9</sup>, such as those that line the gut. Perhaps, from the initial characterization of these channel proteins, Yue *et al.*<sup>2</sup> recognized that they might have properties similar to those of a class of store-operated channels, the  $\text{Ca}^{2+}$ -release-activated  $\text{Ca}^{2+}$  (CRAC) channels<sup>6</sup>, that have been studied intensively by electrophysiological means. Although their molecular components are not known, these channels are among the most selective of known  $\text{Ca}^{2+}$  channels<sup>6</sup>, being able to distinguish  $\text{Ca}^{2+}$  from  $\text{Ba}^{2+}$  ions. CRAC channels also exhibit characteristic behaviours under a variety of experimental conditions. For example, they are regulated by changes in cytoplasmic  $\text{Ca}^{2+}$  levels, and show dramatic changes in selectivity and conductivity in the presence of low extracellular concentrations of divalent cations.

Yue *et al.*<sup>2</sup> have now expressed DNA coding for the CaT1 protein in a mammalian cell line, and show that several of the properties of the encoded proteins closely resemble those of CRAC channels. Notably, as well as there being qualitative similarities between CRAC channels and CaT1, single CaT1 channels conduct monovalent ions in quantitatively similar ways to CRAC channels when there are no divalent cations in the bathing medium.

The authors also show that the expressed CaT1 channels have the signature property of capacitative  $\text{Ca}^{2+}$ -entry channels — they are activated by depletion of intracellular  $\text{Ca}^{2+}$  stores. Interestingly, the authors see this only when lower than maximal amounts of CaT1 are expressed. This may reflect a limiting quantity of other cellular factors needed to regulate the channel. In other words, CaT1 may not be regulated directly by depletion

of intracellular  $\text{Ca}^{2+}$  stores, but rather indirectly, through such cellular factors.

The weight of evidence supports Yue *et al.*'s conclusion that CaT1 (or perhaps, in some instances, its close relative ECaC) constitutes the ion-conducting pore of CRAC channels. But there are also store-operated  $\text{Ca}^{2+}$ -conducting channels, known from electrophysiological studies, that have properties distinct from those of CRAC channels<sup>5</sup>. It is possible, and even likely, that other channel proteins — such as members of the more studied 'short' TRP family<sup>10,11</sup> — form part of these channels. In the near future, I anticipate continuing progress in the search for the complete molecular definition of the capacitative  $\text{Ca}^{2+}$ -entry channels, as well as a solution to the mystery of how they are regulated. ■

James W. Putney Jr is in the Laboratory of Signal

#### Global change

## Time, money and tradeoffs

David F. Bradford

Measurement of the total emissions of greenhouse gases implies specifying 'tradeoffs' of one against another, as in the Kyoto Protocol. In that process, economics has to be taken into account.

In designing policy to check global warming, it is useful to think in terms of an aggregate of various greenhouse gases. The Kyoto Protocol of 1997, for example, uses this idea in prescribing limits on emissions by industrialized countries. The protocol takes in six gases, aggregated into  $\text{CO}_2$  equivalents by applying the concept of 'global warming potential' (GWP); this is a purely physical measure adopted by the Intergovernmental Panel on Climate Change<sup>1</sup> and inspired by a similar concept used in the context of ozone<sup>2</sup>.

In their paper on page 675 of this issue<sup>3</sup>, Manne and Richels bring economics into consideration by use of a model that integrates physical and economic systems. They look at tradeoffs, the principle that a nation can emit more of one gas if it emits less of another, in particular at tradeoffs between methane and nitrous oxide — the two major greenhouse gases other than  $\text{CO}_2$  — and  $\text{CO}_2$ . They show that such tradeoffs meet different policy aims, in terms of intended effect on the climate over a certain time, than those met by GWPs; that they are highly sensitive to the details of the specific aims; and that they are likely to have to vary considerably over time.

Reducing greenhouse-gas emissions can only be done at a cost in goods and services forgone. That cost varies across time according to the gas concerned. There is a continuum of emission trajectories for the several gases that will achieve any given climate

Transduction, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA.

e-mail: putney@niehs.nih.gov

1. Putney, J. W. Jr *Capacitative Calcium Entry* (Landes Biomedical Publishing, Austin, TX, 1997).
2. Yue, L., Peng, J.-B., Hediger, M. A. & Clapham, D. E. *Nature* **410**, 705–709 (2001).
3. Berridge, M. J. *Nature* **361**, 315–325 (1993).
4. Birnbaumer, L. *et al. Proc. Natl Acad. Sci. USA* **93**, 15195–15202 (1996).
5. Putney, J. W. Jr & McKay, R. R. *Bioessays* **21**, 38–46 (1999).
6. Parekh, A. B. & Penner, R. *Physiol. Rev.* **77**, 901–930 (1997).
7. Harteneck, C., Plant, T. D. & Schultz, G. *Trends Neurosci.* **23**, 159–166 (2000).
8. Peng, J.-B. *et al. J. Biol. Chem.* **274**, 22739–22746 (1999).
9. Hoenderop, J. G. J. *et al. J. Biol. Chem.* **274**, 8375–8378 (1999).
10. Boulay, G. *et al. Proc. Natl Acad. Sci. USA* **96**, 14955–14960 (1999).
11. Freichel, M. *et al. Nature Cell Biol.* **3**, 121–127 (2001).
12. Randriamampita, C. & Tsien, R. Y. *Nature* **364**, 809–814 (1993).
13. Irvine, R. F. *FEBS Lett.* **263**, 5–9 (1990).
14. Berridge, M. J. *Biochem. J.* **312**, 1–11 (1995).



In tackling global warning, the economic perspective matters. But does the United States mean business? (See News section.)

objective — that is, there are different means to the same end — so tradeoffs between the gases can critically affect the overall cost.

In regard to their concentrations in the atmosphere, the equivalents among greenhouse gases are relatively straightforward. They all affect radiative forcing: the re-radiation of heat by greenhouse-gas molecules back towards Earth's surface, usually expressed relative to a preindustrial historical baseline. Radiative forcing is transformed by complex feedbacks — for example, cloud formation — into features such as average surface temperature.

In principle, one can calculate the radia-

tive forcing of given concentrations of the six greenhouse gases at a given point in time. And one can calculate the amount of CO<sub>2</sub> in the atmosphere that, taken alone, would result in the same radiative forcing. This would then be the 'CO<sub>2</sub> equivalent' concentration. But it is not obvious how to define CO<sub>2</sub>-equivalent emissions of different greenhouse gases: such emissions affect gas concentrations in the future, as well as today, because each of them has a different lifetime in the atmosphere.

The GWP of a gas is a measure of such an equivalence. The starting point is a baseline trajectory of radiative forcing over time. If, at some moment, an extra unit (say a tonne) of a gas is injected into the atmosphere, the trajectory is perturbed from that moment on. One measure of the resulting warming effect from that moment to some specified time horizon is simply the integral of that perturbation in radiative forcing. The GWP of a gas is the ratio of that integral to the same measure (out to the same horizon) associated with injection of a unit of CO<sub>2</sub>.

In general, natural scientists have been attracted to the GWP concept because of its purely physical quality. Although economists have argued that the tradeoffs cannot be inferred from physical properties alone, but have an inherent economic and policy dimension in terms of targets set<sup>4-7</sup>, the message has been slow to be accepted in the scientific community<sup>8,9</sup>.

Using a computer model that has been used for various purposes for several years, Manne and Richels<sup>3</sup> show both that the economic perspective matters and that it is possible to implement it. They emphasize that there remains much uncertainty about the parameters of their model, and policy objectives. The important point about their results is less the specific tradeoffs that they derive than the cost-effectiveness logic they apply.

From a technical point of view, climate change is simpler than many environmental challenges: here we are dealing with a set of well-mixed pollutants, each of which has the same effect. The policy interest in the emissions of these gases results from their effects on climate change, but this depends on one thing: the trajectory of radiative forcing. That is, we can completely characterize the consequence of an incremental tonne emitted of a greenhouse gas, holding constant all other emission flows through time, by the perturbation in the trajectory of radiative forcing that it induces. The perturbation that results from an incremental tonne varies a lot from gas to gas (and with the ambient atmospheric conditions, although this effect is generally ignored in these exercises).

To say that emission of an incremental tonne of one gas has the same implications for policy as emission of  $x$  tonnes of another gas means that we have to assign a value to the change in radiative forcing at different

times in the future. There is no way to avoid this step. Using the physical GWP involves an implicit evaluation: a bit of extra radiative forcing at any time up to the chosen horizon has the same (negative) value as the same bit at any other time within that span; an extra bit beyond the horizon has zero value. This is clearly wrong.

The grand aim of optimizing policy on climate change involves placing a value on the costs or benefits of such change — say of a 2°C rise in temperature. That is a highly controversial step. But suppose we had somehow solved that grand problem. The solution would involve a particular path of radiative forcing. Presumably, if we have really solved the grand problem, the emissions of the various greenhouse gases minimize the cost of attaining that path. Manne and Richels have recognized that we do not know the overall costs and benefits of climate change: instead of solving the grand problem they have solved a cost-effectiveness subproblem.

They take as given a desired limit on global average temperature, and use their model to solve the cost-minimizing way to achieve the specified target by controlling emissions of the various greenhouse gases. (This involves determining a path of radiative forcing, as well as the cost-minimizing paths of greenhouse-gas emissions.) Associated with a solution to this problem will be 'shadow prices' which answer the question of how much the cost of meeting the desired policy objective would be reduced if we were allowed to emit one extra tonne of gas  $g$  at time  $t$ . The ratio of that quantity for gas  $g$  to that quantity for CO<sub>2</sub> is the tradeoff we are after, and which Manne and Richels calculate.

The significance of Manne and Richels' analysis is twofold. First, they have brought a powerful quantitative model to bear on the problem of aggregating greenhouse gases in the implementation of climate policy. Second, the general point that they emphasize is still far from generally understood. I often encounter the view that, arbitrary as they may be, GWPs constitute a reasonable shot at the set of tradeoffs that are appropriate for Kyoto-style regulation. But this is nonsense without a concept of what 'appropriate' means, which in turn forces us to confront the specifics of getting the tradeoffs right. ■

David F. Bradford is in the Woodrow Wilson School, Princeton University, Princeton, New Jersey 08544-1013, USA.

e-mail: bradford@princeton.edu

1. Schimel, D. et al. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. et al.) (Cambridge Univ. Press, 1996).
2. Lashof, D. A. & Ahuja, D. R. *Nature* **364**, 529–531 (1990).
3. Manne, A. S. & Richels, R. G. *Nature* **410**, 675–677 (2001).
4. Eckaus, R. S. *Energy J.* **13**, 25–35 (1992).
5. Reilly, J. M. & Richards, K. H. *Environ. Res. Econ.* **3**, 41–61 (1993).
6. Schmalensee, R. *Energy J.* **14**, 245–255 (1993).
7. Wallis, M. K. & Lucas, N. J. D. *Int. J. Energy Res.* **18**, 57–62 (1994).
8. Smith, S. J. & Wigley, T. M. L. *Clim. Change* **44**, 445–457 (2000).
9. Smith, S. J. & Wigley, T. M. L. *Clim. Change* **44**, 459–469 (2000).

## Daedalus

### Warping space

Some while ago, Daedalus devised a method of warping space itself. He used a big capacitor, across which an a.c. electric field maintained an a.c. displacement current. All dielectrics maintain such a current, as their electrons are shifted by the electric field; a magnetic field imposed on the dielectric therefore exerts a force upon it. Sadly, a resonant system with the magnetic field created by the same coils that generate the electric one fails to work (the resonance comes out wrong), so the magnetic field has to be specially generated.

The original idea was to use the device as an aero engine, propelling air as the dielectric. But Daedalus soon realized that the system has a remarkable property. Suppose the dielectric is a pure vacuum. This sustains a displacement current with the best of them. Yet the magnetic field, which imposes its force on any current-carrying conductor, is now exerting that force on space itself.

Thus Daedalus's gadget is an ideal way of studying the 'flexibility' of pure space. Cosmology currently holds that the entire Universe is expanding, like a balloon with markings on it being inflated — the markings in this case being galaxies, which all seem to be receding from one another. This expansion may be an energetic relic of the Big Bang, or it may be maintained by some unrecognized force. Daedalus now plans to measure that force.

His scheme is to set up a line of his electromagnetic thrusters, all pushing on space in the same direction, and to shine a laser beam across it. When the thrusters are turned on, the beam (travelling in the space above them) will be deflected. Interferometric measurements on the beam will thus calibrate the thrusters against the resulting beam deflection.

From the results, Daedalus hopes to measure the force sustaining the observed expansion of the Universe, and to relate it to that exerted by the Big Bang. Of course, space-warp technology has already been developed so enthusiastically by science-fiction writers that Daedalus sees no point in entering the business himself. Instead, he hopes that the measurements show extreme sensitivity in the beam. This would imply that the galaxies are expanding purely ballistically, as suggested by the Big Bang theory. But a small force, either positive or negative, may be needed to square an exactly closed Universe with the Big Bang theory as currently understood. Such fundamental measurements are always worth making.

David Jones



# Enigmatic northern plains of Mars

A network of ridges in this region opens a new tectonic window onto this planet.

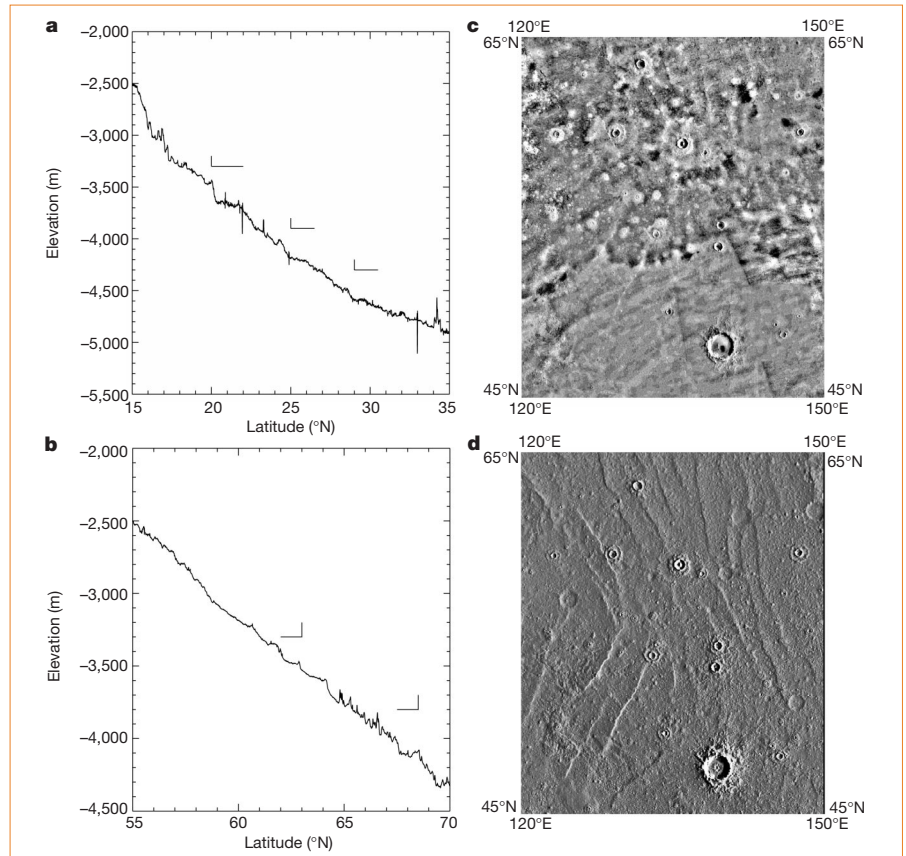
**A**lthough the northern plains of Mars form the flattest known surface in the Solar System, they are crisscrossed by ridge features<sup>1</sup>. Here we test the idea that they might once have been covered by an ocean<sup>2</sup> by examining the topographic profiles of possible shorelines. We conclude that these candidate shorelines were more likely to have been formed by tectonic rather than oceanic processes.

Linear slope changes in the northern plains have been identified as possible shorelines of an ancient ocean formed during the middle period of martian history<sup>2</sup>. Figure 1 shows topographic profiles, generated from Mars Orbiter Laser Altimeter (MOLA) data<sup>1</sup>, across two such groups of shorelines. Candidate shorelines near the Utopia impact basin are flat terraces, with a higher ridge bounding their landward, or upslope, side (Fig. 1a). Possible shorelines on the other side of the proposed ocean, near the Alba Patera volcano, are also flat terraces, with a raised ridge bounding their oceanward, or downslope, side (Fig. 1b).

We believe that this morphology is hard to explain in terms of a shoreline-formation process, as is the reversal of shoreline morphology from one side of the ocean to the other. We favour the idea that these candidate shorelines were created by tectonic activity, on the basis of recent MOLA digital terrain models of kilometre-scale horizontal resolution of the northern plains of Mars.

As seen from early images recorded by Viking (Fig. 1c), these plains are essentially flat and featureless, but MOLA data (Fig. 1d) reveal a network of ridges spanning the northern plains, some of which are the candidate shorelines of the proposed ancient ocean<sup>2</sup>. Most ridges appear to be related to obvious stress centres, such as the volcanic Tharsis Rise, the Utopia impact basin and the Alba Patera volcano. These ridges are generally perpendicular to the predicted directions of maximum compressive stress, which indicates that the ridges have a tectonic origin<sup>3</sup>. They also have the characteristic profile of wrinkle ridges formed by compressive tectonism<sup>4</sup>. Some ridges are close to known wrinkle ridge provinces, such as Lunae Plenum, and have similar strikes; clearly, these formed with the known wrinkle ridges. Both groups of candidate shorelines have orientations consistent with their formation by compressive tectonism.

The causes of the youth and smoothness of the northern plains are still debated. This network of ridges is the only tectonic feature in this region and their discovery opens a new tectonic window onto Mars.



**Figure 1** Martian topography. **a**, MOLA profile 10190 near the Utopia impact basin. Terraces and ridges are marked by horizontal and vertical lines, respectively. **b**, MOLA profile 10929 near the Alba Patera volcano. Terraces and ridges are marked by horizontal and vertical lines, respectively. Vertical exaggeration in **a** and **b** is by a factor of ~400. **c**, Viking photomosaic near the Utopia impact basin, Mercator projection. **d**, Shaded relief map generated from MOLA digital terrain model of the same region with the same projection. Many ridges are visible in the MOLA image that are not evident in the Viking image.

Paul Withers\*, Gregory A. Neumann†‡

\*Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

e-mail: withers@lpl.arizona.edu

†Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡Laboratory for Terrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

1. Smith, D. E. *et al.* *Science* **284**, 1495–1503 (1999).

2. Head, J. W. III *et al.* *Science* **286**, 2134–2137 (1999).

3. Banerdt, W. B. *et al.* in *Mars* (eds Kieffer, H. H. *et al.*) 249–297 (Univ. Arizona Press, Tucson, 1992).

4. Schultz, R. A. *J. Geophys. Res.* **105**, 12035–12052 (2000).

## Food-web dynamics

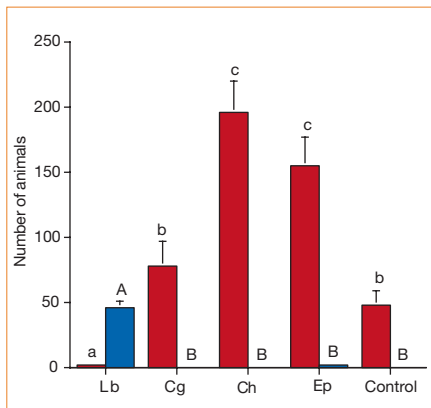
### Animal nitrogen swap for plant carbon

**P**redatory plants are typically found in discrete groups living under conditions of extreme nutrient stress. But we show here how a common species of boreal tree can act indirectly as a predator of arthropods living in the soil by virtue of a fungal symbiont that supplies it with animal nitrogen in exchange for the plant's carbon. If the way in which this partnership operates proves to be widespread, ideas

about nutrient cycling<sup>1,2</sup> and food-web dynamics<sup>3</sup> in temperate forests may have to be modified.

Ectomycorrhizas are mutualisms between soil fungi and the roots of trees that improve plant nutrition, specifically by nitrogen, in exchange for carbohydrates<sup>4</sup>. Soil arthropods such as springtails and mites selectively feed on fungi, including mycorrhizal, decomposer and pathogenic groups<sup>5</sup>.

During a routine feeding study in microcosms, however, we found that less than 5% of springtails (*Folsomia candida*, a soil-dwelling arthropod) survived after two



**Figure 1** Predation on animals by an ectomycorrhizal fungus. Effect of different soil fungi on the survival of springtails is shown, where 'Lb' and 'Cg' represent the ectomycorrhizal fungi *Laccaria bicolor* and *Cenococcum geophilum*, respectively, which are both associated with the tree roots of *Pinus strobus*; Ch, *Cladosporium herbarum* and Ep, *Epicoccum purpurascens*, both fungal saprobes; control, no fungus. All organisms were collected from a forest dominated by *P. strobus* in Ontario, Canada. Experimental units consisted of a single fungus growing on nutrient agar in Petri dishes. Fifty springtails placed on the fungal cultures were incubated for two weeks at 20 °C and the number of living (red bars) and dead (blue bars) springtails then recorded. Animals were sectioned and analysed for internal hyphal infection. Lettering on bars represents differences at the significance level  $P < 0.05$  following a Tukey *post-hoc* test;  $n = 10$ . Error bars represent 1 s.e.

weeks' exposure to the ectomycorrhizal fungus *Laccaria bicolor* (Fig. 1). All dead animals were internally infected by *L. bicolor*. In contrast, when grown with other test fungi, most animals survived and actively reproduced (Fig. 1;  $F = 2.34$ ,  $P = 0.00001$ ). No animals died after the other treatments, including in the fungus-free control (Fig. 1;  $F = 1.97$ ,  $P = 0.00001$ ).

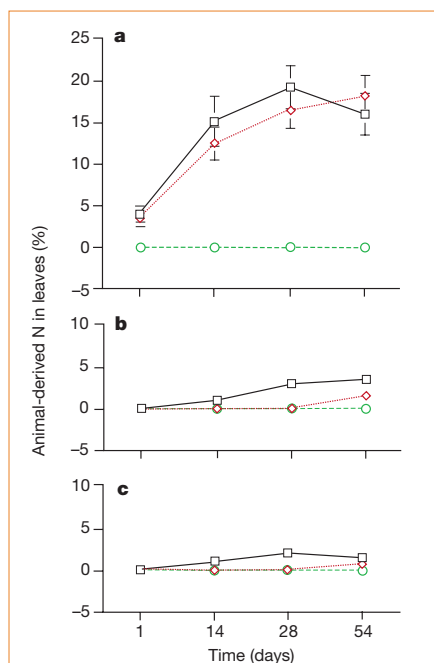
This indicated that *L. bicolor* might be acting as a predator on these arthropods. Careful observation of the microcosms revealed that the ectomycorrhizal fungus immobilized the animals before infecting them. A similar strategy is used by a fungal saprobe, which produces the paralyzing toxin ostreatin<sup>6</sup>. We showed that a proportion of animals were still alive after being immobilized by *L. bicolor* by their positive reaction to a vital stain, fluorescein diacetate; springtails killed in ethanol were unstained. This result indicates that *L. bicolor* may also produce a toxin to paralyse its prey before hyphal invasion.

To determine the fate of the arthropod-derived nitrogen, we used <sup>15</sup>N-labelled

springtails as prey in another microcosm study and determined the amount of <sup>15</sup>N ending up in the foliage of young *Pinus strobus* seedlings reared in the presence of ectomycorrhizal fungi to test whether this animal-origin nitrogen taken up by *L. bicolor* could be transferred to the plant host.

Springtails (alive or already dead) labelled with <sup>15</sup>N were added to the microcosms containing mycorrhizal or non-mycorrhizal plants. The experiment was set up so that only the fungus and not the roots made contact with the animals. The amount of nitrogen in plant tissues and extraradical fungal hyphae of animal origin were determined over a two-month period.

Significant fungus ( $F = 1.15$ ,  $P = 0.0003$ ), animal ( $F = 2.00$ ,  $P = 0.009$ ) and fungus  $\times$  animal ( $F = 1.86$ ,  $P = 0.016$ ) effects were found for <sup>15</sup>N in plant tissues. Up to 25% of plant nitrogen was derived from springtails in the presence of *L. bicolor* (Fig. 2a;  $F = 1.34$ ,  $P = 0.00001$ ), which was able to acquire and transfer nitrogen to its host plant from either living or dead animals with comparable efficiency. At final harvest,



**Figure 2** Percentage of plant nitrogen derived from springtails by an ectomycorrhizal fungus. Results obtained with a, *Laccaria bicolor*; b, *Cenococcum geophilum*; and c, without fungus. Nitrogen source: squares, live springtails; diamonds, dead springtails; and circles, no springtails. Each experimental unit consisted of a  $30 \times 30 \times 1$ -cm glass microcosm containing silica sand, separated into two compartments by a  $30\text{-}\mu\text{m}$  mesh. One pre-germinated *P. strobus* seedling was planted in the left compartment, together with two 10-mm agar plugs of one of the two fungi added at 5 cm below the seedling. After one month of growth, 500 springtails labelled with <sup>15</sup>N were added to the right compartment. Dead springtails were added after being killed by exposure to 50% ethanol for 30 min. Only the fungi were able to penetrate the mesh and move across both compartments. Plants were watered as needed and fertilized weekly with a low-nitrogen-containing solution. The amount of animal-origin nitrogen in the plant tissues was determined using a NOI-5 emissions spectrophotometer. Nitrogen utilization was calculated as  $100 \times (\% \text{ } ^{15}\text{N in seedling leaves or fungal hyphae} / \% \text{ } ^{15}\text{N-labelled springtails})$ . Plant biomass, <sup>15</sup>N in fungal tissues, and the number of active animals remaining were determined on the final day. Results were analysed using a repeated measures factorial ANOVA;  $n = 10$ . Error bars represent 1 s.e. Further details are available from the authors.

plant biomass was stimulated in the *L. bicolor* + springtail treatments (mean, 1.34 g; s.e., 0.16) compared to all other treatments combined (0.81 g; s.e., 0.22) ( $F = 2.02$ ,  $P = 0.00001$ ). Animal extraction at the end of the experiment retrieved less than 10% of the number of animals at the start (mean, 47.2 individuals; s.e., 6.3).

In contrast, seedlings grown with a different ectomycorrhizal fungus (Fig. 2b;  $F = 6.34$ ,  $P = 0.003$ ) or in the absence of fungus (Fig. 2c;  $F = 12.39$ ,  $P = 0.34$ ) were only able to acquire nitrogen from dead springtails (less than 5% of plant N derived from springtails). Animals in these treatments had higher reproductive success and we extracted more animals than were added at the start (mean, 649.1 individuals; s.e., 98.2).

We conclude that ectomycorrhizal plants can indirectly depredate soil arthropods for a significant source of nitrogen through their fungal partners. Our results reveal a nitrogen cycle of far greater flexibility and efficiency than was previously assumed<sup>1</sup>, where the fungal partner uses animal-origin nitrogen to 'barter' for carbon from the host tree. The host, in turn, supplies its fungal associate with carbon to synthesize proteolytic enzymes. Should this phenomenon prove to be widespread, forest-nutrient cycling may turn out to be more complicated and tightly linked than was previously believed.

**John N. Klironomos, Miranda M. Hart**

Department of Botany, University of Guelph, Ontario N1G 2W1, Canada  
e-mail: jklirono@uoguelph.ca

- Schlesinger, W. H. *Biogeochemistry: An Analysis of Global Change* (Academic, San Diego, 1991).
- Simard, S. W. et al. *Nature* **388**, 579–582 (1997).
- Moore, J. C., Walter, D. E. & Hunt, H. W. *Annu. Rev. Entomol.* **33**, 419–439 (1988).
- Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, San Diego, 1997).
- Fitter, A. H. & Sanders, I. R. in *Mycorrhizal Functioning, an Integrative Plant–Fungal Process* (ed. Allen, M. F.) 333–354 (Chapman & Hall, New York, 1992).
- Barron, G. L. & Thorn, R. G. *Can. J. Bot.* **65**, 774–778 (1987).

## Power laws

## Are hospital waiting lists self-regulating?

We have discovered that waiting lists to see hospital consultants are subject to the power laws of complexity theory, and so are likely to be an essential symptom of an efficient healthcare service. Like other complex networks, both privately and publicly funded healthcare systems probably organize themselves in such a way as to reduce the impact of any attempted intervention. They self-regulate to buffer against differing levels of demand, thereby creating bottomless pits that absorb all resources made available. It seems that we would be wiser to judge the system by measuring the overall quality of medical care, rather than by the length of hospital waiting lists.

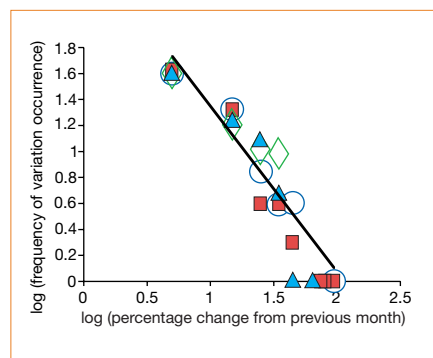


The time taken for patients referred by their general practitioner to see a specialist consultant varies from month to month. We analysed this month-to-month variation in the waiting lists for four dermatology specialists over a 6-year period (data published in monthly reports) using calculations similar to an early analysis of monthly cotton prices<sup>1</sup>. To our surprise, a double-logarithmic plot of the magnitude of monthly variations in waiting time against the frequency with which variations of this magnitude occur gave a straight line, indicating that this month-to-month variation was not random at all (Fig. 1). There is a clear bias towards smaller variations, which occur frequently, with larger variations occurring only rarely; this phenomenon is apparent over a large range. Moreover, because the variations plotted in this way fall on a straight line, they are conforming to what is termed a power law.

All four consultants' waiting lists followed exactly the same pattern, mapping onto the same logarithmic plot, despite their different clinical sub-specializations and workloads (Fig. 1). This similarity (known as finite scaling) is another characteristic feature of complex systems — the individual components are self-similar — implying that the consultants are subject to the same occult agenda. This agenda is not set by the patients, doctors or managers, but is an emergent property of the overall system and cannot be predicted from inspection of the individual interacting parts. Waiting lists in the privately funded sector follow a similar power law (data not shown).

Power laws are used to describe the behaviour of systems that are semi-chaotic, such as sandpile avalanches<sup>2</sup>, forest fires<sup>3</sup> and disease epidemics<sup>4</sup>. Small variations in the magnitude of an event occur frequently, whereas large catastrophic changes are rare. There is buffering against large changes, although the system is not completely immune to them. Complex systems also often represent a perversely efficient system that can self-organize<sup>5</sup> to cope with low levels of resource supply. Such systems endure intervention, like adding more sand to the pile, fighting the fire, or introducing healthcare initiatives to shorten waiting lists, with minimal alteration to their external appearance — so the sand slope remains the same, the fire burns itself out, and the waiting list is unchanged. This is because anything designed to make an impact on such a complex system is dissipated throughout the network of connected participants, so it appears to have a self-regulating life of its own.

Hospital waiting lists therefore represent a complex system composed of a series of interconnected components such as the supply of specialists and the referral rate



**Figure 1** Monthly waiting-list variations are shown for four consultants (represented by different symbols) over a six-year period. The month-to-month variations in waiting time  $t$  are easily calculated as, for example,  $t_{\text{April}} - t_{\text{March}} / t_{\text{April}}$ . A wide spread of month-to-month variations is found and a frequency distribution for each magnitude of variation is shown. The double-logarithmic plot reveals a power law relating frequency and magnitude of variation (Pearson coefficient  $R=0.96$ ,  $P<0.0001$ ).

from general practitioners. Skin diseases are common, but for every affected person who visits their family doctor, there are many more who do not. These people represent a large hidden part of the 'sandpile' of demand for medical care<sup>6</sup> by filling in the gaps when, for example, there are more doctors available or access to them is easier. As the sandpile is much bigger than the part the doctors see, huge changes in supply (on a logarithmic scale) will probably be needed to meet that demand. That demand may then change again, creating an even greater demand, as new diseases appear in society. Simply throwing in extra resources to meet demand<sup>7</sup>, like providing an extra specialist for a few months, may help the waiting list temporarily but, like digging into a dry sandpile, the system then self-organizes around another level of demand until a balance is reached, which is then kept in check by the 'efficient' waiting-list system. If self-organization is occurring, then, the system must be operating at an efficient equilibrium for the level of input.

What are the practical implications of all this? Caution is called for in extending our observations based on one speciality to a single National Health Service teaching hospital to encompass nationwide health-service systems. If it turns out that waiting lists in general conform to power laws, then this could explain why attempts to reduce them have only a temporary effect. A paradoxical feature of waiting lists that conform to a power law is that they represent the most efficient configuration for that organization. Healthcare systems should therefore be judged by measurements taken from their patients and not their waiting lists.

Interconnectivity is an important ingredient of complexity modelling and imposes the potential for catastrophic large-scale events to be brought about by small innocuous ones. If the system is critically self-orga-

nized and not chaotic, these large-scale events are rare. Computer models of waiting lists, which are only semi-chaotic, should provide insight into the main components of the waiting-list phenomenon.

**D. P. Smethurst, H. C. Williams**

Centre of Evidence-Based Dermatology,  
Queen's Medical Centre, University Hospital,  
Nottingham NG7 2UH, UK

e-mail: dominic.smethurst@nottingham.ac.uk

1. Mandelbrot, B. B. *J. Business* **33**, 394–419 (1963).
2. Bak, P. *How Nature Works: The Science of Self-organized Criticality* (Springer, New York, 1996).
3. Morein, G. & Turcotte, D. L. *Science* **281**, 1840–1842 (1998).
4. Rhodes, C. J. & Anderson, R. M. *Nature* **381**, 600–602 (1996).
5. Banavar, J. R., Maritan, A. & Rinaldo, A. *Nature* **399**, 130–132 (1999).
6. Williams, H. C., Stevens, A. & Raferty, J. in *Health Care Needs Assessment* 2nd ser., 261–348 (Radcliffe Medical, Oxford, 1997).
7. Harrison, A. *J. Health Serv. Res. Policy* **5**, 64 (2000).

## Materials science

# The hardest known oxide

A material as hard as diamond or cubic boron nitride has yet to be identified<sup>1–6</sup>, but here we report the discovery of a cotunnite-structured titanium oxide which represents the hardest oxide known. This is a new polymorph of titanium dioxide, where titanium is nine-coordinated to oxygen in the cotunnite (PbCl<sub>2</sub>) structure. The phase is synthesized at pressures above 60 gigapascals (GPa) and temperatures above 1,000 K and is one of the least compressible and hardest polycrystalline materials to be described.

The hardness of ionic and covalent materials is related to their elastic properties and increases with bulk modulus ( $K_T$ ) and shear modulus<sup>1,5,7–9</sup> (Table 1). Titanium dioxide may have a series of high-pressure phases with a hardness approaching that of diamond<sup>10,11</sup>. Cotunnite-structured (OII phase; space group  $Pnma$ ) ZrO<sub>2</sub> and HfO<sub>2</sub> have extremely high bulk moduli of 444 and 340 GPa, respectively<sup>6</sup>. If TiO<sub>2</sub> were able to exist in the cotunnite structure, then this material should also be incompressible and very hard.

We performed lattice-dynamic and *ab initio* LCAO-HF and FPLMTO simulations for plausible structures at pressures up to 100 GPa to identify any that TiO<sub>2</sub> might adopt under high pressure. Structures we simulated were rutile, anatase, TiO<sub>2</sub>-II, baddeleyite-type (MI phase; space group  $P2_1/c$ ), pyrite ( $Pa3$ ), fluorite ( $Fm3m$ ), OI ( $Pbca$ ) and OII (refs 6, 10, 11). The results indicated that cotunnite-structured TiO<sub>2</sub> should be the most stable above 50 GPa (Fig. 1). Our *ab initio* calculations predicted a remarkably high bulk modulus value for this phase:  $380 \pm 20$  GPa by LCAO-HF, and  $386 \pm 10$  GPa by FPLMTO calculation.

We used electrically or laser-heated diamond anvil cells to try and create

**Table 1** Hardness of polycrystalline materials

| Material                          | Bulk modulus (GPa) | Hardness* (GPa) | Ref. |
|-----------------------------------|--------------------|-----------------|------|
| B <sub>4</sub> C                  | 200                | 30 (30)         | 3    |
| SiC                               | 248                | 29 (29)         | 3    |
| Al <sub>2</sub> O <sub>3</sub>    | 252                | 20 (19)         | 7    |
| SiO <sub>2</sub> , stishovite     | 291                | 32 (33)         | 7    |
| WC                                | 421                | 30 (30)         | 8    |
| Cubic BN                          | 369                | (32)            | 3    |
| Cotunnite-type TiO <sub>2</sub> † | 431                | 38              | 3    |
| Sintered diamond                  | 444                | (50)            | 3    |

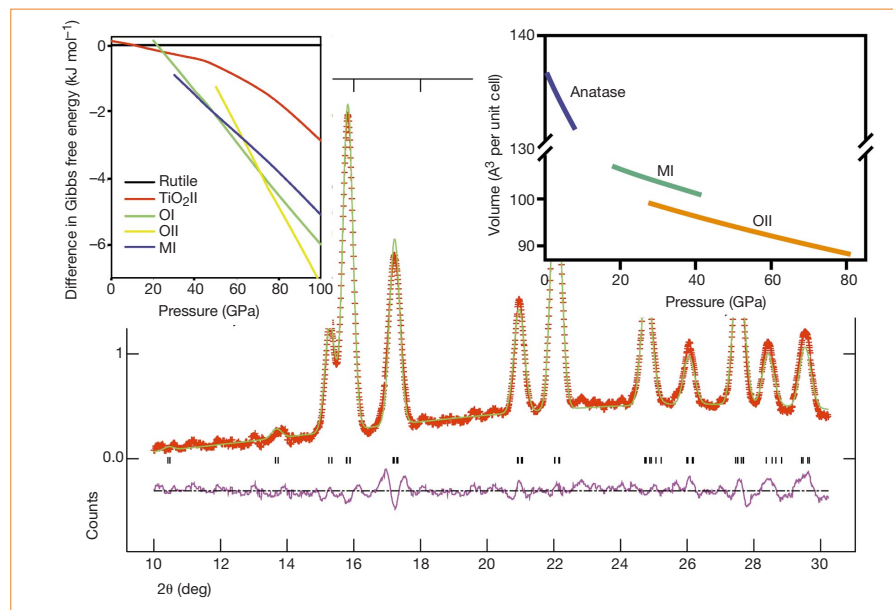
\*Literature data are given in parentheses. The uncertainty in measured hardness is less than 3 GPa.

†Measurements were made at 157 ± 2 K.

cotunnite-structured TiO<sub>2</sub>, with anatase or rutile (99.99% TiO<sub>2</sub>) as starting material. At or over about 12 GPa, both transformed to the baddeleyite structure<sup>10,11</sup> (Fig. 1). On further compression, reflections due to the MI phase could be followed up to over 60 GPa. Unit-cell parameters of the MI phase were determined at 15–42 GPa and molar volume versus pressure data were fitted to a third-order Birch–Murnaghan equation of state (Fig. 1). This gave values for bulk modulus  $K_{300}$  and its pressure derivative,  $K'$ , of  $304 \pm 6$  GPa and  $3.9 \pm 2$ , respectively, in agreement with reported values<sup>11</sup>. Above about 45 GPa the diffraction pattern suffered, and at about 60 GPa the material became translucent.

After heating at 1,600–1,800 K by laser for 40 min at 60–65 GPa, the material transformed into a new phase (Fig. 1). Rietveld refinement of the X-ray powder diffraction data<sup>12</sup> from a sample synthesized in an electrically heated diamond anvil cell at  $61 \pm 2$  GPa and  $1,100 \pm 25$  K (Fig. 1) yielded atomic positions in the  $Pnma$  space group similar to those of PbCl<sub>2</sub> and cotunnite-type ZrO<sub>2</sub> (ref. 6). The transition from MI to OII increases the coordination number of titanium from seven to nine, with the oxygen atoms forming elongated tricapped trigonal prisms containing the titanium atoms.

The cotunnite-type TiO<sub>2</sub> phase could be compressed at ambient temperature to at least 80 GPa, decompressed to below 30 GPa, and then ‘recompressed’ to higher pressures (Fig. 1). Fitting the  $P$ – $V$  data collected at ambient temperature to a third-order Birch–Murnaghan equation of state gave values of  $K_{300} = 431 \pm 10$  GPa,  $K' = 1.35 \pm 0.1$  GPa, and  $V_0 = 15.82 \pm 3$  cm<sup>3</sup> mol<sup>−1</sup>. The bulk modulus of cotunnite-type TiO<sub>2</sub>, when extrapolated to ambient conditions, is only slightly lower than that of diamond. On decompression at ambient temperature to below 25 GPa, OII transformed into the MI phase, which transformed to TiO<sub>2</sub>-II on decompression to 8–12 GPa. Rapid decompression (within 1 s) from 60 GPa to ambient pressure in liquid nitrogen at 77 K (using a cryogenic recovery technique<sup>13</sup>) preserved the cotunnite structure. Upon heating at ambient pressure to 175–180 K, the quenched OII



**Figure 1** Example of profile-fitted X-ray diffraction data obtained from a cotunnite-structured TiO<sub>2</sub> sample (space group  $Pnma$ ,  $Z = 4$ ,  $a = 5.163(2)$  Å,  $b = 2.989(1)$  Å,  $c = 5.966(2)$  Å, Ti (0.264(1); 0.25; 0.110(1)), O1 (0.346(1); 0.25; 0.422(1)), O2 (0.012(2); 0.75; 0.325(1)); numbers in parentheses, s.d.). The sample was synthesized in an electrically heated diamond anvil cell at  $61 \pm 2$  GPa and  $1,100 \pm 25$  K and then temperature-quenched to 290 K. The GSAS program<sup>12</sup> was used in Rietveld refinement<sup>12</sup> ( $wR_p = 1.9\%$ ,  $R_p = 1.6\%$ ,  $\chi^2 = 0.49$ ). Left inset shows the stability of various known and hypothetical TiO<sub>2</sub> polymorphs relative to rutile as a function of pressure, obtained by lattice dynamics at  $T = 300$  K; right inset shows the pressure dependence of the volume for the anatase, baddeleyite and cotunnite phases of TiO<sub>2</sub>. Birch–Murnaghan equations of state are plotted as solid lines, with parameters  $K_{300} = 178 \pm 1$  GPa,  $K' = 4$  (fixed) and  $V_0 = 20.59 \pm 1$  cm<sup>3</sup> mol<sup>−1</sup> for anatase;  $K_{300} = 304 \pm 6$  GPa,  $K' = 3.9 \pm 2$  and  $V_0 = 16.90 \pm 3$  cm<sup>3</sup> mol<sup>−1</sup> for baddeleyite type; and  $K_{300} = 431 \pm 10$  GPa,  $K' = 1.35 \pm 0.1$  and  $V_0 = 15.82 \pm 3$  cm<sup>3</sup> mol<sup>−1</sup> for the cotunnite type (OII) phase.

phase transformed to TiO<sub>2</sub>-II.

We tested cotunnite-structured TiO<sub>2</sub> at 155–160 K with a Vickers microhardness tester (Shimadzu type M) using loads of 25, 50, 100, 150 or 300 g. The hardness of several polycrystalline materials quenched after sintering in a diamond anvil cell at 9–11 GPa and  $770 \pm 25$  K was in good agreement with published results (Table 1).

Cotunnite-structured TiO<sub>2</sub> samples for testing were prepared by heating anatase or rutile to  $1,100 \pm 25$  K at 60–70 GPa in an electrically heated diamond anvil cell for 7 to 8 h. After synthesis, samples of cotunnite-type TiO<sub>2</sub> were cryogenically recovered at 77 K. They were cylindrical (diameter, 250–280 μm; thickness, 40–60 μm), with clean, flat surfaces suitable for hardness measurements. In nine independent indentation measurements of the hardness of cotunnite-type TiO<sub>2</sub>, we found that the hardness was high and independent of load (hardness, 36.8–40.7 GPa; mean, 38 GPa).

To our knowledge, this polycrystalline high-pressure cotunnite-structured phase of titanium dioxide is the hardest oxide so far discovered. It is harder than stishovite and boron oxide and much harder than alumina<sup>3,10</sup>. Polycrystalline cubic boron nitride and sintered diamond are about half as soft as their single-crystal equivalents<sup>3,7</sup>. This suggests that cotunnite-type TiO<sub>2</sub> must be among the hardest known polycrystalline materials.

**L. S. Dubrovinsky\*, N. A. Dubrovinskaya\*, V. Swamy†, J. Muscat†, N. M. Harrison‡, R. Ahuja§, B. Holm§, B. Johansson||**

\*Institute of Earth Sciences, Uppsala University, S-752 36 Uppsala, Sweden

e-mail: leonid.dubrovinsky@geo.uu.se

†CSIRO Minerals, Box 312, Clayton South, Victoria 3169, Australia

‡CCLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD and Department of Chemistry, Imperial College London, London SW7 2AY, UK

§Condensed Matter Theory Group, Department of Physics, Uppsala University, S-751 21 Uppsala, Sweden

||Applied Materials Physics, Department of Materials Science and Engineering, Royal Institute of Technology, Brinellvägen 23, SE-100 44, Stockholm, Sweden

- Cohen, M. L. *Solid State Commun.* **92**, 45–52 (1994).
- Riedel, R. *Adv. Mater.* **6**, 549–556 (1994).
- Sung, C. M. & Sung, M. *Mater. Chem. Phys.* **43**, 1–6 (1996).
- Léger, J. M., Haines, J. & Blanzat, B. *J. Mater. Sci. Lett.* **13**, 1688–1692 (1994).
- Lundin, U. *et al. Phys. Rev. B* **57**, 4979–4982 (1998).
- Desgreniers, S. & Lagarec, K. *Phys. Rev. B* **59**, 8467–8472 (1999).
- Léger, J. M. *Nature* **383**, 401 (1996).
- Teter, D. M. *Mater. Res. Sci. Bull.* **22**, 27 (1998).
- Jhi, S. H. *et al. Nature* **399**, 132–134 (1999).
- Haines, J. & Léger, J. M. *Physica B* **192**, 232–239 (1993).
- Olsen, J. S., Gerward, L. & Jiang, J. Z. *J. Phys. Chem. Solids* **60**, 229–233 (1999).
- Larson, A. C. & Von Dreele, R. B. *Los Alamos National Laboratory, LAUR*, 86 (1994).
- Leinenweber, K., Schuelke, U., Ekundit, S. & McMillan, P. F. in *Properties of Earth and Planetary Materials at High Pressure and Temperature* (eds Manghni, M. H. & Yagi, T.) **101**, 97–105 (Am. Geophys. Union, Geophys. Monogr., Washington, 1998).



# Invariant scaling relations across tree-dominated communities

Brian J. Enquist\* & Karl J. Niklas†

\* Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA; National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, California, USA

† Department of Plant Biology, Cornell University, Ithaca, New York 14853, USA

**Organizing principles are needed to link organismal, community and ecosystem attributes across spatial and temporal scales. Here we extend allometric theory—how attributes of organisms change with variation in their size—and test its predictions against worldwide data sets for forest communities by quantifying the relationships among tree size–frequency distributions, standing biomass, species number and number of individuals per unit area. As predicted, except for the highest latitudes, the number of individuals scales as the  $-2$  power of basal stem diameter or as the  $-3/4$  power of above-ground biomass. Also as predicted, this scaling relationship varies little with species diversity, total standing biomass, latitude and geographic sampling area. A simulation model in which individuals allocate biomass to leaf, stem and reproduction, and compete for space and light obtains features identical to those of a community. In tandem with allometric theory, our results indicate that many macroecological features of communities may emerge from a few allometric principles operating at the level of the individual.**

Despite insights from ecological theory and experimental manipulation, mechanistic connections among important characteristics of ecological communities across diverse ecosystems have remained elusive<sup>1–7</sup>. Variation in species diversity has been explained in terms of a subdivision of community niche space with a presumed concomitant change in total standing biomass and productivity<sup>2,3,8,9</sup>. Yet, the ecological properties of communities or ecosystems accounting for this remain contentious<sup>2,5,10–13</sup>.

One promising mechanistic approach interrelating many organismal, community and ecosystem properties is to focus on size-dependent (allometric) relationships that demonstrably cut across phylogenetically disparate species<sup>6–7,14–25</sup>. One of the most prevalent allometric patterns observed for both plant and animal communities is the inverse relationship between body mass and abundance<sup>23,26,27</sup>. Because this relationship reflects how biomass and productivity are partitioned among individuals, it offers considerable insight into the mechanisms structuring ecological communities across varying environments. However, relatively little is known about how community size–frequency distributions vary across different environments or how they vary among communities differing in species composition. Here we provide a broad theoretical framework for the size–frequency distributions of plant communities. We also show how the allometric constraints on resource use and plant form influence many of the macroscopic properties of tree-dominated communities.

## Extending allometric theory to plant communities

Allometric theory<sup>16</sup> predicts that the total number of individuals,  $N$ , in any size class  $m$ , equals  $C_m M^{-3/4}$ , where  $C_m$  is the number of individuals per unit area normalized to a given size class  $m$ , and  $M$  is the total body mass in class  $m$ . This relationship is predicted to hold true when all available space is occupied such that the total rate of resource use of all individuals within a community  $Q_{\text{Tot}}$  (which is proportional to rates of gross primary production<sup>17</sup>) approximates the rate of resource supply from the environment  $R$  (that is,  $Q_{\text{Tot}} \approx R$ )<sup>16</sup>. Biomechanical and allometric theory<sup>14,15</sup> also predicts that  $M$  is proportional to the  $8/3$  power of stem diameter  $D$  of any size class (that is,  $M \propto D^{8/3}$ ), such that  $N$  will scale as  $N \propto M^{-3/4} \propto D^{-2}$ .

If these scaling laws hold for entire communities, organismal traits can be used to link to larger-scale properties of communities across different ecosystems. For example, extensions of allometric

and biomechanical theory predict that total standing community biomass will be invariant with respect to species composition and thus latitude. Furthermore, the intrinsic capacity to produce biomass on an annual basis will vary little across communities. Note that total standing community biomass,  $M_{\text{Tot}}$ , is given by the formula

$$M_{\text{Tot}} = C_m \int_a^b M^{-3/4} dM = 4C_m [M_a^{1/4} + M_b^{1/4}] \quad (1)$$

where the subscripts  $a$  and  $b$  denote maximum and minimum body mass within a given community, respectively. As both the minimum and maximum body sizes are largely insensitive to species composition or latitude<sup>14</sup> (see also results below), any variation in  $M_{\text{Tot}}$  will be determined by variation in  $C_m$ . For closed canopy forest, however, both theory and observation suggest that  $C_m$  varies little, such that  $M_{\text{Tot}}$  is expected to vary little across communities.

Specifically, for any given size class,  $R_m \approx Q_m \approx C_m B_m$ , where the metabolic rate  $B_m = C_B A_m$ . Here,  $A_m$  is leaf or root area, and  $C_B$  is the rate of resource use per unit area, which can vary across species. Because allometric theory and empirical data<sup>14,16–18</sup> show that  $A_m = C_A (M/\rho)^{3/4}$ , where  $\rho$  is the bulk tissue density and  $C_A$  is a constant of proportionality reflecting the species-specific amount of leaves or roots per individual per unit area, we derive the formula

$$C_m \approx \frac{R_m}{C_A C_B (M_m/\rho)^{3/4}} \quad (2)$$

which quantitatively shows how numerous factors can influence plant population density per size class. Nonetheless, biometric and physiological data indicate no significant differences in the mean values of  $C_B$ ,  $C_A$  and  $\rho$  across tropical and temperate tree species or with variation in species richness<sup>16,17,20–22</sup> (see also Methods). This invariance indicates that total community biomass is likely to be insensitive to species diversity, even though  $C_m$  can vary because of the many environmental factors (such as temperature and precipitation) known to influence  $R_m$ .

Variation in species diversity might also independently influence  $R_m$ , the quantity of resources used per size class per unit time. In particular, variation in  $R_m$  might result either from an increase in total niche volume occupied or as a consequence of synergistic effects from increased diversity such as overyielding<sup>10–11,13</sup>, both of which can influence the total number of individuals across size classes and thus alter total community biomass<sup>4</sup>. A full treatment of

ontogenetic growth is complicated; however, empirical and theoretical data for plants, spanning 20 orders of magnitude in size (unicellular algae to trees), indicate that short-term rates of biomass production per individual,  $G$ , which approximates the whole-organism metabolic rate,  $B$ , scales as  $G \approx B \propto M^{3/4}$  (refs 17, 18, 25). Thus, interspecific rates of biomass production are predicted to be intrinsically equivalent in relation to size within and across plant communities.

As the total rate of community resource use  $R_{\text{Tot}} \propto NQ \propto NB \approx G_{\text{Tot}}$ , where  $G_{\text{Tot}}$  is the net primary production of biomass of a given community, it follows that  $G_{\text{Tot}} \approx QN \propto M^{3/4}M^{-3/4} \propto M^0$  (ref. 16). This equation predicts that rates of production both within and across communities are invariant with plant size. Furthermore, if  $C_m$ ,  $M_a$  and  $M_b$  (from equation (1)) do not vary across communities, then it also follows that variation in rates of plant community total biomass production,  $G_{\text{Tot}}$ , are more influenced by ecological factors that reduce the capacity of metabolic production<sup>6</sup> (for example, abiotic and biotic features of ecosystems which influence the extent to which plants can maximally transpire water and assimilate  $\text{CO}_2$ ) than by species-specific physiological capacities or variation in species diversity.

### Empirical validation of allometric theory

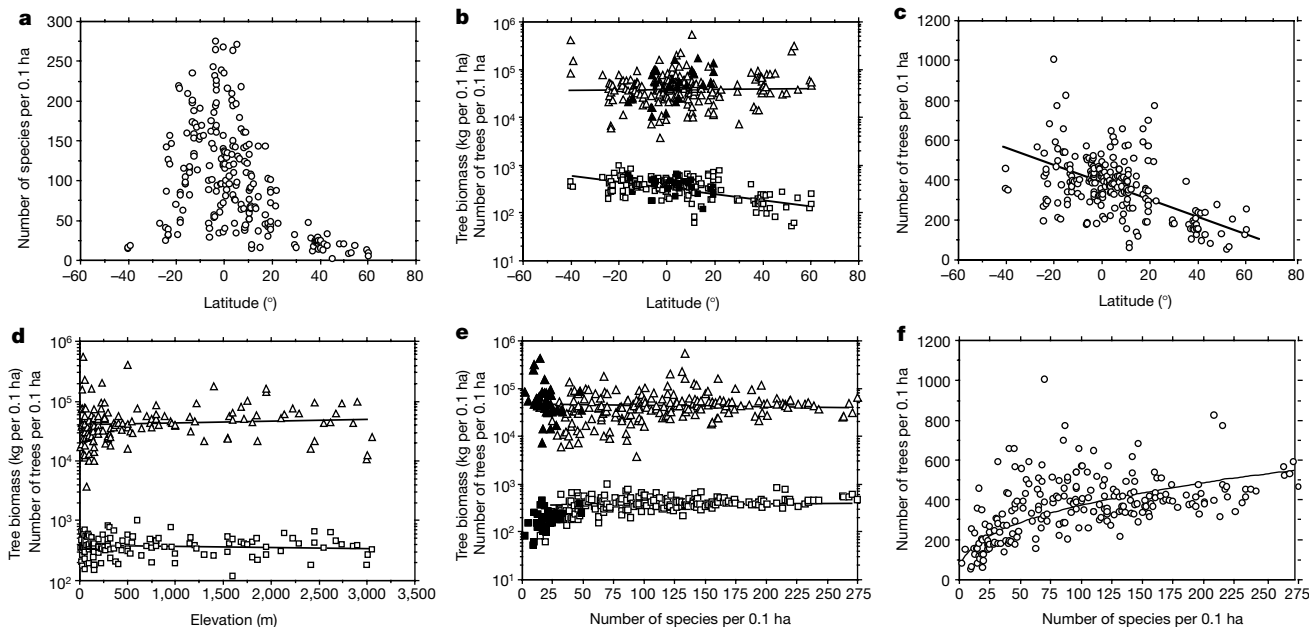
Clearly, these predictions, which follow directly from biophysical first principles dictating size-dependent optimal biomass allocation patterns<sup>14–17</sup>, require empirical verification. Here we test these predictions on the basis of macroecological data sets that span taxonomically and physiognomically diverse plant communities. We primarily draw on a large data set assembled by Gentry<sup>28,29</sup>, which spans near-monospecific stands to some of the most bio-

diverse forested communities on Earth. Specifically, the Gentry data set represents a 22-year accumulation from 227 sites across 6 continents of tropical and temperate closed-canopy forest communities ranging from 60.4°N to 40.43°S latitude and from 20 to 3,050 m in elevation.

In each site sampled, all plants, including lianas with stem diameters  $\approx 2.5$  cm measured at breast height (d.b.h.), were sampled along ten 2 × 50-m transects, totalling 0.1 ha at each site. Species numbers and numbers of individuals per site are in the ranges 2–275 and 52–1,005, respectively. The complete data set contains 83,121 individual plants (maximum d.b.h. ranges between 26 and 412 cm). (Information and access to the data are available at [http://www.mobot.org/MOBOT/Research/applied\\_research/gentry.html](http://www.mobot.org/MOBOT/Research/applied_research/gentry.html); see also Supplementary Information.)

### Results

On the basis of the above protocol, the maximum number of species per 0.1 ha increases toward the equator (Fig. 1a). Yet, total tree standing biomass per 0.1 ha is invariant with respect to species number, latitude or elevation, even though tree density increases from northern to southern latitudes (Fig. 1b–d). Likewise, species diversity has no effect on total standing biomass, although tree density rapidly asymptotes with respect to species diversity (Fig. 1e–f). Finally, as predicted by theory, the number of individuals per sample area scales as the  $-2$  power of stem diameter or as the  $-3/4$  power of body mass both within and across communities (Fig. 2). It seems as though both temperate and tropical forests are described by similar size distributions (Fig. 2a). Notably, the  $-2$  scaling rule also holds with increasing geographic sampling areas, including continental and global samples (Fig. 2). Latitude and



**Figure 1** Properties of tree-dominated communities. Solid lines are regression curves based on model type I regression analyses. **a**, Number of tree species per 0.1 ha plotted against latitude. **b**, Biomass (triangles) and number of trees per 0.1 ha (squares) plotted against latitude (negative values indicate degrees south) (regression of untransformed data for tree biomass and number against untransformed latitude:  $n = 220$ ,  $r^2 = 0.0024$ ,  $F = 0.5283$ ,  $P = 0.4681$  and  $n = 220$ ,  $r^2 = 0.2570$ ,  $F = 77.49$ ,  $P < 0.0001$ , respectively). Filled symbols denote data for communities from high elevations ( $\approx 1,200$  m), all of which are from low latitudes ( $\leq 20^\circ$  S and N). **c**, Number of trees per 0.1 ha plotted against latitude (see **b** for log-arithmetic plot of untransformed data). **d**, Number of trees per 0.1 ha (squares) and total standing tree biomass (triangles) plotted against elevation (regression of untransformed data against elevation gives

$n = 145$ ,  $r^2 = 0.0007$ ,  $F = 0.1063$ ,  $P = 0.7449$  and  $n = 141$ ,  $r^2 = 0.0004$ ,  $F = 0.052$ ,  $P = 0.820$ , respectively). **e**, Number of trees (squares) and total standing tree biomass (triangles) per 0.1 ha plotted against number of species per 0.1 ha (regression of log-transformed data for tree number against log-transformed data for species number per 0.1 ha gives  $n = 227$ ,  $r^2 = 0.4789$ ,  $F = 206.8$ ,  $P < 0.0001$ , see **f**; regression of untransformed data for tree biomass per 0.1 ha against species number gives  $n = 221$ ,  $r^2 = 0.0013$ ,  $F = 0.294$ ,  $P = 0.588$ ). Closed symbols denote data from high elevations ( $\approx 1,200$  m) or high latitudes (all high elevation plots are from low latitudes, that is,  $\leq 20^\circ$  S and N; see **b**). **f**, Number of trees per 0.1 ha plotted against species number per 0.1 ha (see **e** for log-arithmetic plot).



species number do not contribute greatly to the variance observed in local size distribution exponents ( $n = 226$ ,  $F = 50.05$ ,  $r^2 = 0.183$ ,  $P < 0.0001$  and  $n = 200$ ,  $F = 22.76$ ,  $r^2 = 0.092$ ,  $P < 0.0001$ , respectively) (Fig. 3a, b). Furthermore, neither the size–frequency distribution exponent nor the number of individuals is correlated with annual precipitation<sup>28,29</sup> ( $n = 91$ ,  $F = 0.893$ ,  $r^2 = 0.0099$ ,  $P < 0.347$ , and  $n = 91$ ,  $F = 2.545$ ,  $r^2 = 0.0278$ ,  $P < 0.1142$ , respectively).

Neither latitude nor elevation can serve as a surrogate measure of climate. Here, each is used simply as an independent variable against which to plot total tree standing biomass across communities differing in species composition as well as geography<sup>28,29</sup>. Likewise, the biomass reported here neither include trees smaller than 2.5-cm d.b.h. nor that contained in roots or nonarborescent vegetation (see Methods). For each community, however, the largest trees per site account for between 15% and 27% of the total standing tree biomass, whereas maximum tree size is invariant with respect to latitude ( $n = 225$ ,  $r^2 = 0.00001$ ,  $F = 0.0002$ ,  $P = 0.9896$ ) or species number ( $n = 226$ ,  $r^2 = 0.008$ ,  $F = 1.755$ ,  $P = 0.187$ ).

The principal exceptions in our findings are the size distributions for communities from high latitudes ( $\geq 40^\circ$  N or S). For these communities, the size distribution exponent tends to be less negative than  $-2$ , indicating lower densities of smaller individuals (see Figs 2 and 3). However, no significant correlation is observed between the exponent of the size frequency distribution and elevation ( $n = 144$ ,  $r^2 = 0.015$ ,  $F = 2.179$ ,  $P = 0.1421$ ). Clearly, one or more ecological features (for example, periodic recruitment, disturbance and/or length of the growing season) probably further limits the ability of individuals at the highest latitudes to fully occupy physical space. Yet even for these communities total standing biomass is, on average, indistinguishable from that of communities from lower latitudes or elevations, suggesting compensation by the largest individuals. Furthermore, with

increased geographic sampling area, high-latitude North American forests approximate the  $-2$  scaling rule (see Fig. 1, legend).

The available data for tree-dominated communities do not permit a direct test of our prediction that the rate of resource use and mass production in tree communities is invariant with respect to body size or species composition<sup>17</sup>. However, as (1) both total biomass and size–frequency distributions vary little with respect to latitude or tree species composition, and (2) a single allometric function describes interspecific rates of biomass production across all plants<sup>17,18</sup>, strongly suggesting this invariance across natural stands, the intrinsic capacities to use resources and to produce biomass per unit area may be equivalent.

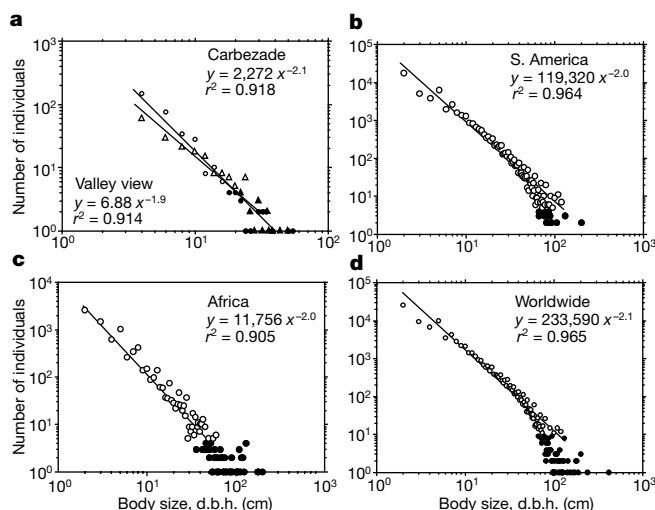
Variance in resource use and ultimately annualized rates of net primary production (NPP) across closed-canopy forest communities reflects the influence of many different ecological factors that reduce the capacity of metabolic production (for example, abiotic and biotic features of ecosystems that influence the extent to which plants can maximally transpire water and assimilate  $\text{CO}_2$ ; ref. 6) rather than variation in species diversity. This prediction is in accord with recent findings for North American herbaceous communities showing little relationship between empirical measures of NPP and plant species richness<sup>12</sup>. Similarly, gross primary production (of which NPP is an approximate constant fraction<sup>3,30</sup>) across European forests has been found to be invariant with latitude (a surrogate measure of species richness)<sup>31</sup>.

### An algorithmic model for community dynamics

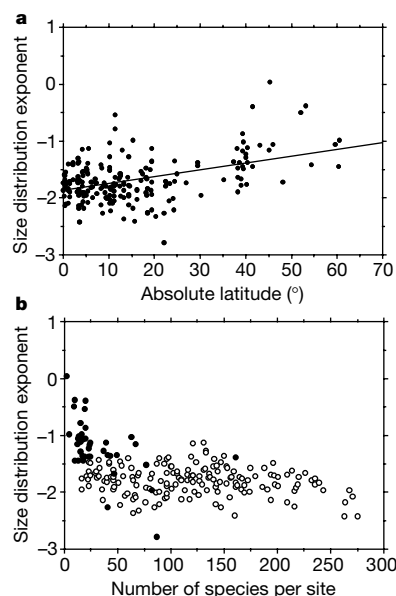
Likewise, when allometric theory is cast in terms of a simple algorithm for biomass allocation at the level of an individual plant, computer models predict that resource use will be invariant within and across monotypic and mixed tree communities. Moreover, all of the observed scaling relationships observed for real tree-dominated communities are predicted when the theory is cast in terms of a computer simulation model.

The conceptual basis for simulating the properties of real communities based on the allocation patterns of individual plants begins with a balanced energy formula

$$E_T - E_r = E_R + E_L + E_S \quad (3)$$



**Figure 2** Size–frequency distributions of forest communities differing in geographic sampling area<sup>8,9</sup> calculated for all size bin classes containing  $\approx 5$  individuals (open symbol) owing to the data sparsity of statistically under-represented size classes (filled symbols). The 95% CI of the exponent for each distribution includes the predicted value of  $-2.0$  based on model type I or II regression analyses. **a**, Overlapping size–frequency distributions of a 0.1-ha sample of the South American tropical forest community located at Cabezeade ( $-10.2^\circ$  latitude; data shown as open circles) and a 0.1-ha sample of a North American community located at Valley View Glades, Missouri ( $38.15^\circ$ ; data shown as open triangles). **b**, Pooled data for all South American forest 0.1-ha sample sites. **c**, Pooled data for all 0.1-ha samples of African forest sites (pooled continental data for the forested communities from Asia and North America also have exponents indistinguishable from  $-2$ ). **d**, Pooled data for all forest 0.1-ha samples represented in the data set (worldwide).



**Figure 3** Exponents of size–frequency distributions of forest communities. **a**, Exponents of each community plotted against absolute latitude. **b**, Exponents of size–frequency distribution of individual communities plotted against species number in each community (filled circles denote data from communities in high latitudes or elevations).

where  $E_T$  denotes the total light energy intake of an individual plant,  $E_r$  designates the energy consumed by respiration, and  $E_R$ ,  $E_L$  and  $E_S$  respectively denote the amount of  $E_T$  available for the construction of reproductive organs, leaves and stems, which differs across species comprising a particular community. In turn, the energy allocated to each of the three above-ground 'compartments' is converted into a biomass specified by the formula

$$aM_T^\alpha = bM_R^\beta + cM_L^\chi + dM_S^\delta \quad (4)$$

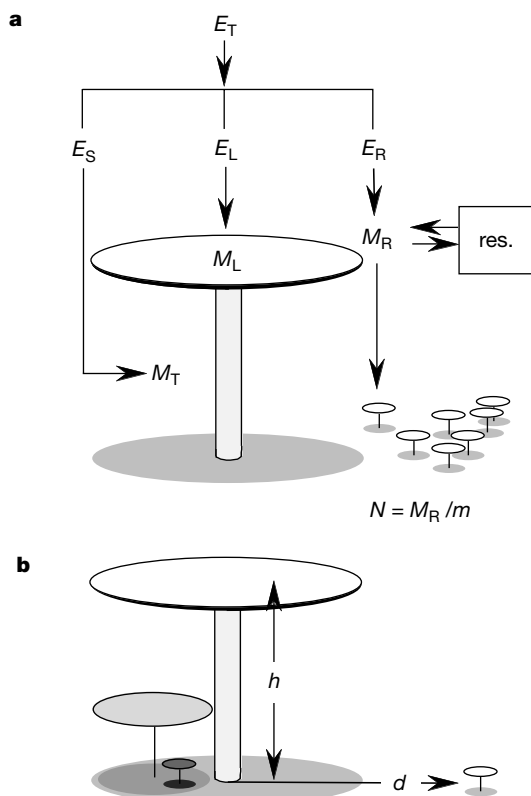
where  $a$ ,  $b$ ,  $c$  and  $d$  are constants of proportionality, and  $\alpha$ ,  $\beta$ ,  $\chi$  and  $\delta$  are species-specific scaling exponents among which the numerical value of  $\alpha$  is dependent on the numerical values assigned to  $\beta$ ,  $\chi$  and  $\delta$ . Within the model, species differ in values of  $a$ ,  $b$ ,  $c$  and  $d$ . For the purposes of this paper, theoretically predicted quarter-power exponents were used for all species<sup>14</sup>. Together, the magnitudes of  $b$ ,  $c$  and  $d$ , and  $\beta$ ,  $\chi$  and  $\delta$  determine biomass allocation patterns across species.

The balanced energy and mass formulas dictating allometric allocation patterns at the level of the individual (see equations (3) and (4)) were incorporated into a computer program designed to predict the consequences of competition for light and space. Each

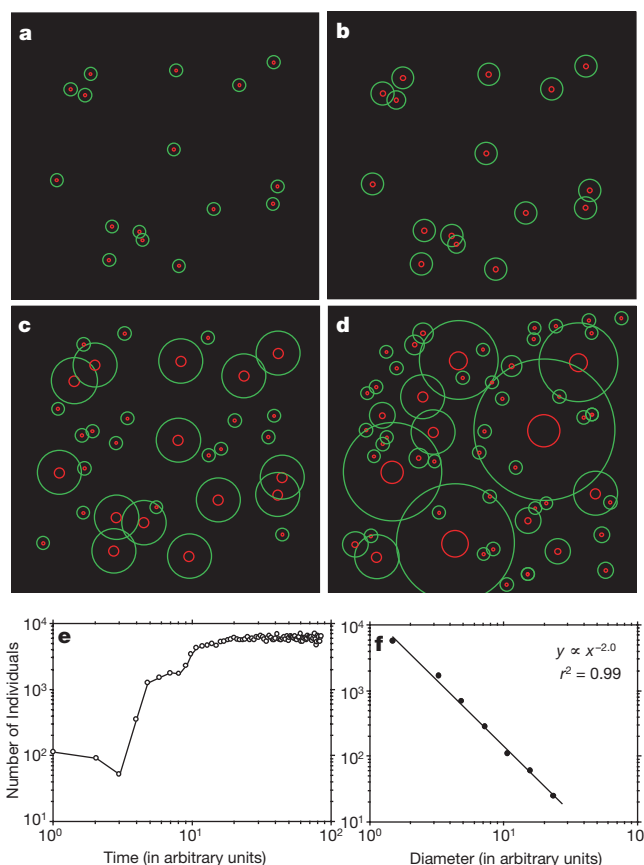
individual consists of a single vertical 'trunk' with height  $h$  and diameter  $d$ , which elevates a circular 'canopy' with radius  $r$  and height  $h$ . The height  $h$  of each individual is a function of the biomass allocation to the stem compartment. In each simulation (see below), the biomass allocated for the construction of leaves and stems is added to each 'growing' individual at the end of each time interval (Fig. 5).

All plants compete for spatially limiting 'light' resources on a world space, which is divided into a system of subgrids. The light energy available to each individual is calculated at each time interval. Each unobstructed canopy element receives 100% sunlight. Individuals located beneath the portions of successively overlapping (shorter) canopies receive proportionately less light energy (expressed as a percentage) defined by a predetermined canopy attenuation factor. This attenuation factor is equivalent for conspecifics but may in principle differ across species. Elements located over trunks receive no light energy.

Depending on its species assignment, an individual may use all or some of its metabolic production for the construction of its reproductive organs during each iterative time interval. In the former case, the total biomass of the  $n$  number of propagules produced at the end of each time interval equals  $M_R$ ; in the case of individuals reserving a portion of their  $M_R$  for subsequent use (that is, those that 'reserve' some or all of their reproductive effort), the total biomass of the  $n$  number of propagules produced at any



**Figure 4** Assumptions used to model biomass allocation among three above-ground plant compartments and stochastic features affecting dispersal and mortality. **a**, Species-specific assumptions. Total energy intake  $E_T$  is partitioned among three compartments (stem  $E_S$ , leaf  $E_L$  and reproductive effort  $E_R$ ) and converted per unit time into biomass (stem  $M_S$ , leaf  $M_L$  and reproductive effort  $M_R$ ) according to specified allometric 'rules'. Some of the available reproductive biomass may be reserved for future use (res.); however, the total number of propagules produced at any time  $N$  is the quotient of the available reproductive biomass and the species-specific biomass of each propagule  $m$  (that is,  $N = M_R/m$ ). **b**, Dispersal and mortality assumptions: propagules are dispersed randomly to a maximum distance  $d$  defined by the height of the parent individual  $h$ . A propagule that begins growth under the canopy of one or more mature individuals receives less light, and, depending on its species-specific minimum light requirement and the species-specific light attenuation of the canopy, either dies or grows less vigorously.



**Figure 5** Results of representative simulations based on the assumptions of the model for biomass allocation (see Fig. 4). **a–d**, Partial polar views of four stages in the 'ontogeny' of a community, from the initial random 'seeding' of propagules (**a**) to the appearance of a 'mature' community consisting of a more or less constant number of individuals (**d**). Large green circles denote canopy radius; small red circles denote stem girth. **e**, Number of individuals plotted against time during the 'ontogeny' of a simulated community. **f**, Size-frequency distribution of a simulated community (number of individuals plotted against stem diameter) with a  $-2.0$  exponent calculated for all size bin classes (see Fig. 2).



time equals some portion of the biomass reserved over all previous time intervals.

A below-ground (root) compartment is not explicitly included in equations (3) and (4), largely because of the many uncertain assumptions required to emulate its metabolic demands, growth and energy-storage capacity. Conceptually, however, an equivalent fraction of the energy/biomass devoted to the construction of each of the three above-ground compartments can be thought of as diverted to the root compartment such that this 'sink' is implicitly treated and can be thought of as subsumed by the variable  $E_r$  in equation (3).

Among 'juveniles', death from light starvation is the result of extensive canopy overlap (that is, the probability of dying is determined by the light attenuation factor and the number of overlapping canopies above shorter individuals). In addition, for all individuals, there is a constant probability of death each time step independently of plant size (this probability can be specified, for example, as a stochastic process, or as a linearly or nonlinearly increasing probability defined as some function of the size of an individual). Seed dispersal is assumed to be passive and modelled according to the ballistic model<sup>32</sup>

$$\frac{d}{h} = \frac{u_s}{u_w} \quad (5)$$

where  $d$  is dispersal distance,  $u_s$  is the settling velocity of the seed (scaled as a function of its biomass, which in turn depends on the total biomass allocated to reproductive effort and how this biomass is subdivided among  $n$  number of seeds), and  $u_w$  is ambient wind speed measured at  $h$  (scaled according to different wind-speed profile formulas used to evaluate their effects). Seed size does not vary within species but can be made to vary across species.

When the allometry of biomass allocation at the level of an individual plant is expressed in terms of theoretically predicted quarter-power scaling functions<sup>14</sup>, the emulated 'ontogeny' of monospecific or mixed plant communities (Fig. 5a–d) obtains a typical sigmoid curve (Fig. 5e) and the scaling relationships  $N \propto D^{-2}$  and  $N \propto M^{-3/4}$  once the size–frequency distribution of a community reaches equilibrium (Fig. 5f). These scaling relationships become more statistically robust as the size (spatial sampling area) of the community is increased. Furthermore, the model predicts equivalent amounts of biomass across communities differing in species composition at equilibrium. These predictions are consistent with findings that the spatial scale of sampling influences the form of the relationship among the number of species per unit area, the number of individuals per unit area, and productivity<sup>12</sup>. Noting that the behaviour of simulated plant communities reflects a biomass optimization process operating at the level of individual plants, we conclude that the invariant properties identified for real plant communities emerge from the allometric rules that influence the behaviour of individual plants competing for space and limited resources.

### Ecological and evolutionary implications

Numerous studies show surprisingly little variation in the basic allometric relationships that define how different tree species occupy space, produce biomass, or apportion biomass among their body parts<sup>6,15,17,20–22,33</sup>. Therefore it is not surprising that our empirical and theoretical findings accord well with published data showing considerable numerical overlap in basal stem area, leaf area and carbon content per unit area among temperate and tropical forests<sup>3,5,28,34,35</sup>. For example, our analysis of additional larger compiled databases<sup>34</sup> indicates that total standing tree biomass (kg dry weight per 0.1 ha) varies little with respect to latitude across communities ( $r^2 = 0.028$ ,  $n = 827$ ,  $F = 23.81$ ,  $P < 0.0001$ ). Also, the total cross-sectional basal area of trees per 0.1 ha, which reflects the capacity of individuals to fill space, varies little with respect to latitude ( $r^2 = 0.029$ ,  $n = 715$ ,  $F = 20.96$ ,  $P < 0.0001$ ).

The assumptions of our allometric model and computer simulation are applicable theoretically to all vascular plant species<sup>36</sup>. With the exception of a small number of parasitic and hemiparasitic species, all tracheophytes compete for light and/or space. Likewise, all share the same basic body plan and must abide by the same general biophysical principles and processes<sup>36,37</sup>. Furthermore, across a broad sampling of diverse plant taxa (unicellular algae, pteridophytes, gymnosperms and angiosperms) biomass production is described by a single allometric relationship<sup>18</sup>. Theoretical plant morphology also suggests that land plant evolution was profoundly influenced by optimization processes required by the trade-offs necessitated by simultaneously performing manifold tasks essential for growth, survival and reproductive success. These trade-offs confined phenotypic expression to a finite number of morphologies<sup>38</sup>. Much evidence thus converges on the conclusion that communities dominated by plants behave similarly by virtue of shared organizing principles operating at the level of the individual.

A growing body of allometric and biomechanical evidence also strongly indicates that plant species have evolved diverse biomass allocation patterns that nonetheless abide by the same fundamental trade-off patterns. Indeed, when seen from an allometric perspective, many of the life-history differences across species reflect subtle (albeit biologically important) variants on a general theme whose mathematical structure is defined by the unavoidable necessity to reconcile conflicting functional design specifications<sup>14,36</sup>. For example, different species allocate different amounts of their annual metabolic production to reproductive effort. Yet, regardless of any particular metabolic or physiological allocation pattern, stoichiometric and allometric relationships influence the nature of biological trade-offs<sup>15–17,39,40</sup>. Trade-offs associated with allometric relationships constrain biomass devoted to reproductive effort and to vegetative growth, as a finite allometrically determined amount of metabolic production is available at any time<sup>15–17</sup>. By altering the timing of biological events, fundamental trade-offs facilitate the coexistence of several species and ultimately result in substantial variations in growth rate, lifespan and reproductive effort<sup>17</sup>.

### Discussion

We have shown that despite wide variation in species diversity, abundance and biomass, tree-dominated communities are characterized by nearly identical size–frequency distributions reflecting nearly equivalent standing biomass. We have also shown that the number of individuals in each community sample scales as  $D^{-2}$  and thus  $M^{-3/4}$ . These observations are consistent with allometric theory and with our computer model of biomass allocation but contrast in many important ways with past speculations and niche-based theoretical predictions<sup>4,10,11</sup>. Allometric theory suggests that variation in plant species composition is instead associated with concomitant changes in the degree of partitioning of a limited amount of resources rather than increases (or decreases) in community biomass and, potentially, depending on the local environment, productivity<sup>4,10–11</sup>. Such partitioning is most probably reflected in life-history trade-offs in the allocation of metabolic production<sup>17</sup>.

Extensions of a general allometric framework and simulation model reveal how several prominent organismal, community and ecosystem level properties emerge from relatively few allometric and biomechanical 'rules'. The constraints of resource transport through 'fractal-like' vascular networks ultimately dictate how individuals fill space, use resources and produce and allocate biomass<sup>14–18</sup>. Such constraints are reflected in allometric scaling relationships, which are evident at many levels in biology. These 'rules' dictate how metabolic production and biomass are partitioned among different body parts at the level of the individual plant. Furthermore, they provide a quantitative basis for drawing mechanistic connections between numerous features of organismal biology, ecology, ecosystem studies and evolutionary biology. □

## Methods

### Calculation of community biomass

To test our theory, we estimated the total standing above-ground biomass. We also determined the size (stem diameter) frequency distribution for all individuals from each site. The total standing (above-ground) tree biomass for each community was calculated from d.b.h. values using empirically determined scaling exponents and constants of proportionality reported in the literature for representative communities growing in latitudes and elevations equivalent to those communities in the Gentry data<sup>22</sup>. Data for trees of tropical dry, moist and wet forest indicate that the interspecific allometry is  $M = 0.124D^{2.53}$ , where  $M$  is measured in kg of dry (above-ground) biomass and  $D$  is in cm d.b.h. (ref. 22). The allometric proportionality constants and the exponents reported in the literature did not significantly vary between temperate and tropical forests<sup>15–17,20–22,33</sup>. Our allometrically estimated community biomass for both temperate and tropical sites per 0.1 is statistically indistinguishable from empirically determined biomass values for forests with similar minimal size cutoffs<sup>33,34</sup>. Nonetheless, we used different scaling exponents to evaluate their impact on estimates of total standing tree biomass. Our analysis showed that using different scaling exponents to estimate total standing biomass had no effect on the invariance of this parameter with respect to latitude or species number. This is because differences in estimated total community biomass remained constant across communities regardless of which scaling exponent was used (that is, the absolute magnitude of standing biomass was affected but not the relative differences across communities).

### Community size distributions

Size class-bins were defined at 2-cm d.b.h. intervals and were then log-transformed. Each size–frequency distribution was characterized in terms of its  $y$  intercept and slope (scaling exponent) using both model type I and model type II regression analyses on log-transformed data. Regression analysis between the slopes obtained using model type I and II regression analyses indicates no significant difference in the scaling exponent ( $r^2 = 0.9728$ ). Model type I regression analysis was used when error estimates in  $x$  values were likely to be minimal (such as latitude, elevation and basal stem diameter (d.b.h.)). Model type II regression analysis was used to determine the allometric scaling exponents for variables neither of which could be assumed to be independent of one another (such as species number and total number of individuals per community)<sup>15</sup>. In most cases, because of high  $R^2$  values, the choice of which regression model has little or no effect on the scaling exponent. Size distribution exponents were calculated for sites where the tree size range exceeded 45-cm d.b.h.

### Computer simulations

Simulations of plant community growth and dynamics were written in objective C and conducted on the SWARM simulation platform based in part at the Santa Fe Institute and developed by the Swarm Development Group. The basic architecture of Swarm is the simulation of collections of concurrently interacting agents (see <http://www.swarm.org/>).

Each simulation begins with a pre-selected number of 'seeds' randomly dispersed in a world-space of predetermined size. The seeds are either all conspecifics (to emulate the behaviour of a monotypic community), or are individually assigned a species-specific biomass, light requirement and dispersal range (to construct a mixed community). In either case, the ambient light intensity is uniform and invariant across all simulations (that is, all unobstructed surfaces receive 100% light energy).

The height  $h$  of each individual is a function of the biomass allocation to the stem compartment. Within the model height cannot exceed the critical Euler buckling height  $h_{crit}$ , which is proportional to the  $2/3$  power of stem diameter; specifically,  $h_{crit} = C(E/\rho)^{1/3}D^{2/3}$ , where  $C$  is a proportionality constant,  $E$  is Young's modulus, and  $\rho$  is average stem-tissue density<sup>15</sup>.

Once a simulation is initiated, all species-specific parameters remain constant (that is, there is no genetic variation among conspecifics and thus no 'evolution'). The time-interval designating the conversion of light energy into biomass is uniform and invariant among all simulations. Community properties such as the number of individuals, total standing biomass, size–frequency distributions (measured in terms of plant height, stem diameter, total biomass, and so on), and other statistical parameters can be retrieved at any time-interval without terminating the simulation.

Received 13 November; accepted 19 December 2000.

1. Brown, J. H. *Macroecology* (Univ. Chicago Press, Chicago, 1995).
2. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
3. Whittaker, R. H. *Communities and Ecosystems* 2nd edn (Macmillan, New York, 1975).
4. Tilman, D. & Pacala, S. in *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds Ricklefs, R. E. & Schluter, D.) 13–25 (Univ. Chicago Press, Chicago, 1993).
5. Huston, M. A. *Biological Diversity* (Cambridge Univ. Press, Cambridge, 1994).
6. Whittaker, R. J. Scaling, energetics and diversity. *Nature* **401**, 865–866 (1999).
7. Ritchie, M. E. & Olff, H. Spatial scaling laws yield a synthetic theory of biodiversity. *Nature* **400**, 557–560 (1999).
8. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton University Press, Princeton, 1967).

9. Terborgh, J. On the notion of favorableness in plant ecology. *Am. Nat.* **107**, 481–450 (1973).
10. Tilman, D. Causes, consequences and ethics of biodiversity. *Nature* **405**, 208–211 (2000).
11. Tilman, D., Lehman, C. L., & Thomson, K. T. Plant diversity and ecosystem productivity: Theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
12. Gross, K. L., Willig, M. R., Gough, L., Inouye, R. & Cox, S. B. Patterns of species diversity and productivity at different scales in herbaceous plant communities. *Oikos* **89**, 417–427 (2000).
13. Huston, M. A. *et al.* No consistent effect of plant diversity on productivity. *Science* **289**, 1255a (<http://www.sciencemag.org/cgi/reprint/289/5483/1255a>) (2000).
14. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664–667 (1999).
15. Niklas, K. J. *Plant Allometry* (Univ. Chicago Press, Chicago, 1994).
16. Enquist, B. J., Brown, J. H. & West, G. B. Allometric scaling of plant energetics and population density. *Nature* **395**, 163–165 (1998).
17. Enquist, B. J., West, G. B., Charnov, E. L. & Brown, J. H. Allometric scaling of production and life-history variation in vascular plants. *Nature* **401**, 907–911 (1999).
18. Niklas, K. J. & Enquist, B. J. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proc. Natl Acad. Sci. USA* **98**, 2922–2927 (2001).
19. Roy, K., Jablonski, D. & Martien, K. K. Invariant size–frequency distributions along a latitudinal gradient in marine bivalves. *Proc. Natl Acad. Sci. USA* **97**, 13150–13155 (2000).
20. Whittaker, R. H. & Woodwell, G. M. Dimension and production relations of trees and shrubs in the Brookhaven Forest, New York. *J. Ecol.* **56**, 1–25 (1968).
21. White, J. The allometric interpretation of the self thinning rule. *J. Theor. Biol.* **89**, 475–500 (1981).
22. Brown, S. Estimating biomass and biomass change of tropical forests. *Food and Agricultural Organization of the United Nations Forestry Paper* 134 (Rome, 1997).
23. Morse, D. R., Lawton, J. H., Dodson, M. M. & Williamson, M. H. Fractal dimension of vegetation and the distribution of arthropod body lengths. *Nature* **314**, 731–733 (1985).
24. Damuth, J. Population density and body size in mammals. *Nature* **290**, 699–700 (1981).
25. Niklas, K. J. Size-dependent variations in plant growth rates and the "3/4-power rule". *Am. J. Bot.* **81**, 134–144 (1994).
26. Barenblatt, G. I. & Monin, A. S. Similarity principles for the biology of pelagic animals. *Proc. Natl Acad. Sci. USA* **80**, 3540–3542 (1983).
27. De Liocourt, F. De l'aménagement des sapinières (*Bulletin de la Société Forestière de Franche-Comte et Belfort*, Besançon, 1898).
28. Gentry, A. H. Changes in plant community diversity and floristic composition on environmental and geographic gradients. *Ann. Miss. Bot. Gar.* **75**, 1–34 (1988).
29. Gentry, A. H. in *Biodiversity and Conservation of Neotropical and Montane Forests* (eds Churchill, S. P. *et al.*) 103–126 (New York Botanical Garden, New York, 1993).
30. Waring, R. H., Landsberg, J. & Williams, M. Net primary production of forests: A constant fraction of gross primary production? *Tree Physiol.* **18**, 129–134 (1998).
31. Valentini, R. *et al.* Respiration as the main determinant of carbon balance in European forests. *Nature* **404**, 861–865 (2000).
32. Okubo, A. & Levin, S. A. A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology* **70**, 329–338 (1989).
33. Jack, S. B. & Long, J. N. Linkages between silviculture and ecology: an analysis of density management diagrams. *Forest Ecol. Manag.* **86**, 205–220 (1996).
34. Cannell, M. G. R. *World Forest Biomass and Primary Production Data* (Academic, New York, 1982).
35. Olson, J. S., Watts, J. A. & Allison, L. J. *Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: A Database CDIC Numeric Data Collection NDP-01 7* (Oak Ridge National Laboratory, Oak Ridge, 1985).
36. Niklas, K. J. *The Evolutionary Biology of Plants* (Univ. Chicago Press, Chicago, 1997).
37. Niklas, K. J. The evolution of plant body plans—a biomechanical perspective. *Ann. Bot.* **85**, 411–438 (2000).
38. Niklas, K. J. Morphological evolution through complex domains of fitness. *Proc. Natl Acad. Sci. USA* **91**, 6772–6779 (1994).
39. Reich, P. B., Walters, M. B. & Ellsworth, D. S. From tropics to tundra: Global convergence in plant functioning. *Proc. Natl Acad. Sci. USA* **94**, 13730–13734 (1997).
40. Elser, J. J. *et al.* Biological stoichiometry from genes to ecosystems. *Ecol. Lett.* **3**, 540–550 (2000).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

### Acknowledgements

We thank D. Ackerly, J. Ackerman, J. Brown, W. Chaloner, K. Harms, J. Haskell, D. Post, T. Phillips, P. Stoll, J. Valentine, R. Whittaker, J. Williams and M. Willig for discussions or comments on the manuscript; J. Miller, P. Raven and the Missouri Botanical Garden for help in disseminating the Gentry database; and J. Pringle and A. Landcaster for programming assistance with SWARM. This work stems in part from the Body Size in Ecology and Evolution Working Group sponsored by The National Center for Ecological Analysis and Synthesis (NCEAS, a centre funded by the NSF, the University of California Santa Barbara and the State of California). In particular we wish to acknowledge J. Haskell for critical help with the Gentry data; G. West and B. Tiffney for support; and F. Smith and the members of the body size working group for their input to this project. B.J.E. was supported by the NSF, the Santa Fe Institute and NCEAS. K.J.N. was supported by an Alexander von Humboldt Forschungspreis and NYS Hatch grant funds.

Correspondence and requests for materials should be addressed to B.E. (e-mail:benquist@u.arizona.edu).



# Stable methane hydrate above 2 GPa and the source of Titan's atmospheric methane

J. S. Loveday\*, R. J. Nelmes\*, M. Guthrie\*, S. A. Belmonte\*, D. R. Allan\*, D. D. Klug†, J. S. Tse† & Y. P. Handa†

\* Department of Physics and Astronomy, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK

† Steacie Institute for Molecular Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

Methane hydrate is thought to have been the dominant methane-containing phase in the nebula from which Saturn, Uranus, Neptune and their major moons formed<sup>1</sup>. It accordingly plays an important role in formation models of Titan, Saturn's largest moon. Current understanding<sup>1,2</sup> assumes that methane hydrate dissociates into ice and free methane in the pressure range 1–2 GPa (10–20 kbar), consistent with some theoretical<sup>3</sup> and experimental<sup>4,5</sup> studies. But such pressure-induced dissociation would have led to the early loss of methane from Titan's interior to its atmosphere, where it would rapidly have been destroyed by photochemical processes<sup>6,7</sup>. This is difficult to reconcile with the observed presence of significant amounts of methane in Titan's present atmosphere. Here we report neutron and synchrotron X-ray diffraction studies that determine the thermodynamic behaviour of methane hydrate at pressures up to 10 GPa. We find structural transitions at about 1 and 2 GPa to new hydrate phases which remain stable to at least 10 GPa. This implies that the methane in the primordial core of Titan remained in stable hydrate phases throughout differentiation, eventually forming a layer of methane clathrate approximately 100 km thick within the ice mantle. This layer is a plausible source for the continuing replenishment of Titan's atmospheric methane.

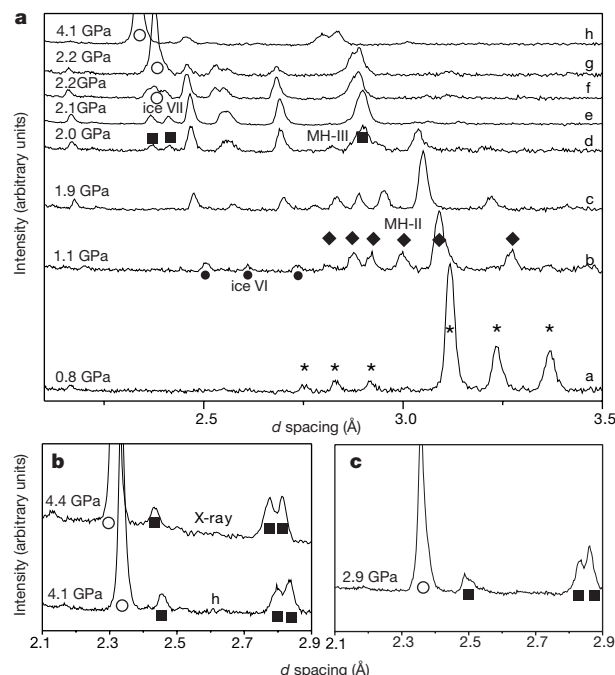
The low-pressure structure of methane hydrate is composed of water cages surrounding methane molecules which stabilize the cages through van der Waals forces—in the 'clathrate I' type of structure<sup>8</sup>. This is referred to as structure-I methane hydrate, or SIMH, and has an ideal composition of 5.75 water molecules to each methane molecule. The stability of clathrate hydrates as a function of pressure has been explored<sup>3</sup> using calculations based on the van der Waals–Platteeuw model<sup>9,10</sup>. These models are general and predict that the clathrate structure becomes unstable with respect to ice and the guest species at the intersection of the decomposition curve with the ice freezing line, because of the greater changes in entropy and volume in forming crystalline clathrate from ice instead of liquid water. In the case of SIMH this is estimated to occur at about 1.2 GPa (ref. 3).

Recent synchrotron X-ray diffraction studies<sup>4,5</sup> have shown a change in structure above 1 GPa, interpreted as the progressive emptying of the smaller cage sites in SIMH—thus liberating free methane—up to 2.2 GPa, where dissociation to ice VII and methane I occurs. This suggests that the expected dissociation occurs as a gradual process over ~1 GPa. However, the fit of this model<sup>4,5</sup> to the observed data is poor. Hirai *et al.*<sup>4</sup> also noted that their finding of three-phase coexistence at room temperature was thermodynamically inconsistent with measurements of the decomposition curve<sup>11</sup>, which identified a triple point at about 0.6 GPa and 320 K, and they speculated that this inconsistency was evidence of metastability. We have now carried out a detailed neutron diffraction study of the behaviour in the range up to 10 GPa on fully deuterated (D-) methane clathrate, and have made comparative synchrotron X-ray studies of H-methane hydrate.

Figure 1a shows neutron diffraction patterns obtained from the

D-hydrate on increasing pressure. Up to 0.8 GPa (profile a), the diffraction pattern was that of SIMH. At 1.1 GPa (profile b), the three strongest peaks of ice VI appeared between 2.5 and 2.75 Å (marked by filled circles), and the formation of a new methane hydrate phase (methane hydrate II, MH-II) was indicated by the six peaks marked by diamonds. On further compression, MH-II underwent a transformation at 2.0 GPa (profile d) to methane hydrate III (MH-III), indicated by the appearance of the peaks marked by squares, with an accompanying increase in height of the ice VI peaks which shows MH-III to be richer in methane than MH-II. At 2.2 GPa the ice VI progressively transformed to ice VII (profiles f and g). MH-III remained stable to 10 GPa without further phase transformation. To mimic the pressure–temperature (*P*–*T*) path followed by the primordial core during Titan's accretion<sup>1</sup>, samples were also compressed to approximately 3 GPa at 130 K and then warmed to room temperature. The sample was found to amorphize on compression and then transform on warming to MH-III and ice VII as can be seen in Fig. 1c.

The observation of new hydrate structures appears to conflict with the behaviour described by refs 4 and 5, and so X-ray data were collected to explore the behaviour of H-hydrate. The sequence of changes was found to be the same as in the D-hydrate up to



**Figure 1** Diffraction patterns showing the three phases of methane hydrate. Peaks are marked for SIMH (asterisks), MH-II (diamonds), ice VI (filled circles), MH-III (squares) and ice VII (open circles). **a**, The neutron diffraction patterns of deuterated methane hydrate on increasing pressure at room temperature. **b**, A comparison of X-ray data collected from H-methane hydrate at 4.4 GPa with neutron data collected from D-hydrate at 4.1 GPa (profile h in **a**). **c**, A neutron diffraction pattern collected from D-methane hydrate at about 3 GPa after being compressed at around 130 K and warmed to room temperature. The samples of H- and D-SIMH were produced by adsorbing methane into ice at –20 °C as described elsewhere<sup>21</sup> and loaded at about 100 K into a Paris–Edinburgh pressure cell<sup>22</sup> for the neutron studies and a Merrill–Bassett diamond-anvil cell for the X-ray studies. The cells were sealed by applying pressure and warmed to room temperature. Fits to diffraction data confirmed full occupancy of the SIMH cage sites. Neutron diffraction studies were carried out on the PEARL beamline<sup>23</sup> of the UK neutron facility, ISIS, at the Rutherford Appleton Laboratory, and X-ray studies were carried out using image-plate techniques<sup>24</sup> on station 9.1 at the UK synchrotron, SRS, at Daresbury Laboratory and beamline ID9 of the European Synchrotron Radiation Facility. Pressures were determined from the known compressibilities of SIMH, ice VI and ice VII in the case of the neutron experiments and using the ruby fluorescence method in the X-ray case.

1.9 GPa—except that MH-II formed at the slightly lower pressure of about 0.9 GPa where water is still liquid<sup>12</sup> and this water superpressed so that ice VI only appeared above 1.6 GPa. (This superpressing of water was also observed in refs 4 and 5, and appears to be a reproducible kinetic effect.) Different behaviours were observed above 1.9 GPa, depending on the rate of compression. A sample taken up in pressure ‘normally’, in steps of some 0.2–0.5 GPa every 1–2 h, transformed as predicted into ice VII and methane I. But the results with the D-hydrate suggested the probable influence of kinetics, and so we held another sample at 1.9 GPa for 12 h. This transformed into a mixture of ice VI and MH-III. At 2.2 GPa the ice VI transformed to ice VII, and at 4.4 GPa this sample gave the pattern shown in Fig. 1b, marked X-ray (profile h of the D-hydrate at a similar pressure is shown for comparison). As for the D-hydrate, hydrogenous MH-III was found to be stable indefinitely up to at least 10 GPa, and samples compressed to approximately 3 GPa at 130 K were found to be MH-III and ice VII when warmed to room temperature.

This behaviour is fully consistent with earlier work. The tabulated *d* spacings from the pattern interpreted as low-occupancy SIMH and ice VI at 1.5 GPa in ref. 5 match the MH-II and ice VI peak positions found in our data, and the (cubic) SIMH structure gives a very poor fit with differences in observed and calculated peak positions as large as 0.2 Å. The transition was found below 1 GPa with solid ice VI not appearing until a significantly higher pressure (1.5 GPa in refs 4 and 5). And no solid methane appeared until the MH-II phase starts to transform to ice and methane<sup>4,5</sup>, as we also observe for the normal rate of increase of pressure. The transition to MH-II is also consistent with the decomposition curve determined in ref. 11 and it is very probable that the “second hydrate” proposed<sup>11</sup> above 0.6 GPa was in fact MH-II.

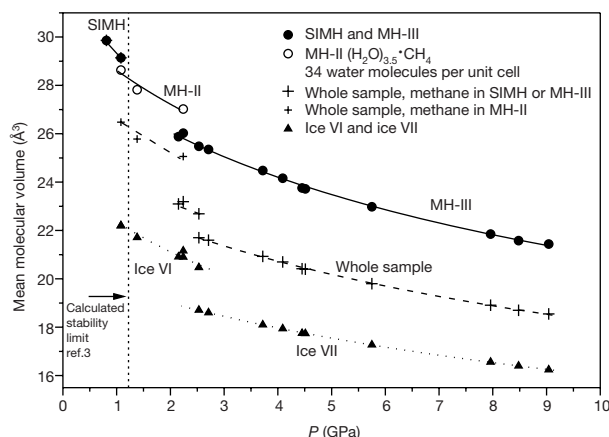
In addition to the three MH-III peaks marked in Fig. 1b, 13 other peaks from MH-III were clearly visible in the X-ray data. From these it was possible to solve the structure as body-centred orthorhombic with unit-cell dimensions  $a = 4.7458(5)$  Å,  $b = 8.0644(9)$  Å,  $c = 7.8453(7)$  Å and  $v = 300.26(3)$  Å<sup>3</sup> at 3.0(1) GPa. The structure has eight water molecules and four methane molecules in the unit

cell, consistent with the composition estimated from the relative peak heights of ice VI and SIMH. The details of this structure solution will be published elsewhere. A Murnaghan fit to the volume as a function of pressure, as shown in Fig. 2, gives an ambient-pressure bulk modulus  $B_0 = 15.2(5)$  GPa with a rate of change with pressure  $B' = 4$ . These results and those for SIMH place constraints on the density of MH-II. On the basis of peak heights, the molecular composition is estimated to be 3.5 water : 1 methane and the patterns can be indexed with a hexagonal unit cell with  $a = 11.7911(12)$  Å,  $c = 9.9210(10)$  Å and  $v = 1194(3)$  Å<sup>3</sup> at 1.7(1) GPa. The volume of the whole sample must decrease with pressure and MH-II must then contain between 28 water molecules with a water : methane ratio of 3 : 1 and 39 water molecules with a ratio of 4 : 1. The values plotted in Fig. 2 are for 34 water molecules and a 3.5 : 1 ratio. This composition is close to the mean value and is also the composition of the hexagonal structure-H (sH) clathrate<sup>13</sup> which has an axial ratio close to that of MH-II. However, although MH-II and sH also have similar pair correlation functions, a fit to the sH structure gave implausible nearest-neighbour oxygen to oxygen distances. We conclude that MH-II does not have the hexagonal clathrate structure but may be closely related to it. (After this work was completed another study observed the same hexagonal phase and concluded on the basis of the  $c/a$  ratio that it has the sH structure<sup>14</sup>. We have rechecked our analysis and have not altered our conclusion.)

These high-pressure structures of methane hydrate have probable relevance to gas hydrates in general. High-pressure transitions have been reported in the hydrates of neon<sup>15</sup>, argon<sup>16,17</sup> and nitrogen<sup>18</sup> and the dissociation curves of neon and argon hydrates<sup>15,16</sup> show similarities to that of methane hydrate<sup>11</sup>. However, none of the differential thermal analysis and light-scattering studies reported until now conclusively indicates the existence of new hydrate structures, as against a change in guest occupancy or in state of ordering<sup>18</sup>. Given the similarities in the dissociation curves, and the fact that nitrogen and argon have similar radii to methane, it is possible that at least argon and nitrogen hydrates also adopt the MH-II structure and that the second transition reported in argon hydrate<sup>17</sup> may be to the MH-III structure.

The discovery that methane hydrates remain stable up to 10 GPa has important consequences for the modelling of Titan. Established models<sup>1,2,19</sup> suggest that at the end of accretion Titan had a primordial core made up of a mixture of (1) rock, (2) ammonia monohydrate, and (3) methane clathrate that had been dissociated into ice and free methane by the core pressures of 2–6 GPa, all surrounded by a thick rocky ‘carapace’. Outside the carapace was a layer of high-pressure ices capped by an ammonia–water ocean, with an outermost surface layer of ice Ih. It is believed that melting in the core through radiogenic heating would have caused eventual rupture of the carapace, leading to core overturn, and gravitational differentiation in which the (low-density) free methane would rise quite rapidly to the surface and form part of Titan’s early atmosphere<sup>1,2</sup>. There is a long-standing debate<sup>6,7</sup> on how the atmospheric methane is replenished, as photochemical processes are expected to destroy the methane too rapidly for the current atmosphere to be a remnant of a primordial methane-rich atmosphere<sup>6,7</sup>. Possible proposed sources include the presence of pools of liquid methane on the surface<sup>6,7</sup>, and a near-surface regolith formed by the interaction of atmospheric or other free methane with surface ice<sup>20</sup>. These depend on an extensive escape of free methane following differentiation, which would require passage of most of the methane through the ammonia–water ocean. Models vary significantly in the predicted nature and thickness of this ocean, depending on assumptions about the composition of the primordial ices, thermal evolution, and so on; the most recent work concludes that the ocean thickness must be too great for methane to traverse without re-enclathration<sup>2</sup>.

Our results change the basis of this debate by showing that all or



**Figure 2** The compression of methane hydrate. The pressure dependence of the mean molecular volume for the three methane-hydrate phases (circles), ices VI and VII (triangles), and the whole sample (crosses)—which is a water–methane mixture corresponding to the 5.75:1 ratio of the initial SIMH—showing that the compression of the overall sample is largely achieved by the transfer of water from methane hydrate phases to the relatively dense ice phases. The values for the whole sample were obtained by placing all the methane and the appropriate amount of water (5.75 water molecules per methane for SIMH, 3.5 water molecules per methane for MH-II, and 2 water molecules per methane for MH-III) in the relevant methane hydrate phase, and the remaining water in the relevant ice phase. The values for MH-II were calculated for the middle of the range of possible values as described in the text. The solid line through the MH-III data is the best-fitting Murnaghan equation of state. All the other lines are guides to the eye.

most of the core methane will remain within the satellite after differentiation, irrespective of the thickness (or existence<sup>19</sup>) of the ocean. First, we show that the material originally accreted into the primordial core as SIMH will be a mixture of ice and MH-III under the  $P$ - $T$  conditions at the end of accretion. As we have also established that the SIMH to MH-II and MH-II to MH-III transitions are both readily reversible, gravitational differentiation after core overturn will return all the core methane to SIMH as the methane-containing component rises above the final core and passes through the layer of ice phases at the bottom of the mantle (which are denser, as Fig. 2 shows). The outcome will be a layer of SIMH some 100 km thick either just below or just above the ammonia-water ocean—the ocean's density is not well enough known to determine which. In either case, this large amount of SIMH will be stable to the base of the surface ice layer (where the pressure is estimated to be 0.1 GPa (ref. 2)).

These results imply that it is unlikely that Titan has surface pools of methane. Instead, the SIMH layer provides a possible long-term source of atmospheric methane through convective processes such as cryovolcanism<sup>1,2,20</sup>. The presence of an approximately 100 km thick SIMH layer will also have an impact on the modelling of thermal and rheological processes<sup>2</sup>. Furthermore, the incorporation into the SIMH layer of a substantial amount of water that was previously believed to be free will increase the ammonia content of the ammonia-water ocean and hence increase its thickness. □

Received 27 June 2000; accepted 26 January 2001.

- Lunine, J. I. & Stevenson, D. J. Clathrate and ammonia hydrates at high-pressure—application to the origin of methane on Titan. *Icarus* **70**, 61–77 (1987).
- Grasset, O., Sotin, C. & Deschamps, F. On the internal structure and dynamics of Titan. *Planet. Space Sci.* **48**, 617–636 (2000).
- Lunine, J. I. & Stevenson, D. J. Thermodynamics of clathrate hydrate at low and high pressures with application to the outer solar system. *Astrophys. J. Suppl. Ser.* **58**, 493–531 (1985).
- Hirai, H. *et al.* Methane hydrate behaviour under high pressure. *J. Phys. Chem. B* **104**, 1429–1433 (2000).
- Hirai, H. *et al.* Methane hydrate, amoeba or a sponge made of water molecules. *Chem. Phys. Lett.* **325**, 490–498 (2000).
- Lorenz, R. D. The weather on Titan. *Science* **290**, 467–468 (2000).
- Lunine, J. I. Does Titan have an ocean? A review of current understanding of Titan's surface. *Rev. Geophys.* **31**, 133–149 (1993).
- Gutt, C. *et al.* The structure of deuterated methane hydrate. *J. Chem. Phys.* **113**, 4713–4721 (2000).
- van der Waals, J. H. Statistical mechanics of clathrate compounds. *Trans. Faraday Soc.* **52**, 184 (1956).
- van der Waals, J. H. & Platteeuw, J. C. Clathrate solutions. *Adv. Chem. Phys.* **2**, 1 (1959).
- Dyadin, Y. A., Aladko, E. Y. & Larionov, E. G. Decomposition of methane hydrates up to 15 kbar. *Mendelev Commun.* **7**, 34–35 (1997).
- Pistorius, C. W., Pistorius, M. C., Blakey, J. P. & Admiraal, L. J. Melting curve of ice VII to 200 kbar. *J. Chem. Phys.* **38**, 600 (1963).
- Udachin, K. A., Ratcliffe, C. I., Enright, G. D. & Ripmeester, J. A. Structure H Hydrate: a single crystal diffraction study of 2,2 dimethylpentane.5(Xe,H<sub>2</sub>S).34H<sub>2</sub>O. *Supramol. Chem.* **8**, 173–176 (1997).
- Chou, I. M. *et al.* Transformations in methane hydrates. *Proc. Natl Acad. Sci.* **97**, 13484–13487 (2000).
- Dyadin, Y. A. *et al.* Clathrate hydrates of hydrogen and neon. *Mendelev Commun.* **5**, 209–210 (1999).
- Dyadin, Y. A., Larionov, E. G., Mirinski, D. S., Mikina, T. V. & Starostina, L. I. Clathrate formation in the Ar-H<sub>2</sub>O system under pressures up to 15,000 bar. *Mendelev Commun.* **7**, 32–34 (1997).
- Lotz, H. T. & Schouten, J. A. Clathrate hydrates in the system H<sub>2</sub>O-Ar at pressures and temperatures up to 30 kbar and 140 °C. *J. Chem. Phys.* **111**, 10242–10247 (1999).
- van Hinsberg, M. G. E., Scheerboom, M. I. M. & Schouten, J. A. The vibrational spectra of N<sub>2</sub> in clathrate-hydrates: A new high-pressure phase transition. *J. Chem. Phys.* **99**, 752–754 (1993).
- Grasset, O. & Sotin, C. The cooling rate of a liquid shell in Titan's interior. *Icarus* **123**, 101–112 (1996).
- Stevenson, D. J. in *Proceedings of Symposium on Titan 29–33* (European Space Agency Special Publication SP-338, Toulouse, 1992).
- Handa, Y. P. Compositions, enthalpies of dissociation, and heat-capacities in the range 85K–270K for clathrate hydrates of methane, ethane, and propane, and enthalpy of dissociation of isobutane hydrate, as determined by a heat-flow calorimeter. *J. Chem. Therm.* **18**, 915–921 (1986).
- Nelmes, R. J. *et al.* Structure studies at high pressure using neutron powder diffraction. *Trans. Am. Cryst. Ass.* **29**, 19–27 (1993).
- PEARL—Pressure and Engineering Research Line 28–29 ISIS'97 Rutherford Appleton Laboratory Report RAL-TR-97-050 (RAL, Didcot, 1997).
- Nelmes, R. J. & McMahon, M. I. High-pressure powder diffraction on synchrotron sources. *J. Synch. Rad.* **1**, 69–73 (1994).

## Acknowledgements

We thank D. J. Francis, T. Bovornratanaraks and M. Hanfland for assistance with experiments, and O. Grasset and D. J. Stevenson for helpful discussions of their work. Our work is funded by a research grant from the EPSRC, and supported by CCLRC through access to beamtime and other resources.

Correspondence and requests for materials should be addressed to J. S. L. (e-mail: J.Loveday@ed.ac.uk).

# A thermodynamic connection to the fragility of glass-forming liquids

L.-M. Martinez & C. A. Angell

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA

Although liquids normally crystallize on cooling, there are members of all liquid types (including molecular, ionic and metallic) that supercool and then solidify at their glass transition temperature,  $T_g$ . This continuous solidification process exhibits great diversity within each class of liquid—both in the steepness of the viscosity-temperature profile, and in the rate at which the excess entropy of the liquid over the crystalline phase changes as  $T_g$  is approached. However, the source of the diversity is unknown. The viscosity and associated relaxation time behaviour have been classified between 'strong' and 'fragile' extremes, using  $T_g$  as a scaling parameter<sup>1</sup>, but attempts to correlate such kinetic properties with the thermodynamic behaviour have been controversial<sup>2,3</sup>. Here we show that the kinetic fragility can be correlated with a scaled quantity representing excess entropy, using data over the entire fragility range and embracing liquids of all classes. The excess entropy used in our correlation contains both configurational and vibration-related contributions. In order to reconcile our correlation with existing theory and simulations, we propose that variations in the fragility of liquids originate in differences between their vibrational heat capacities, harmonic and anharmonic, which we interpret in terms of an energy landscape. The differences evidently relate to behaviour of low-energy modes near and below the boson peak.

The understanding of fluidity and diffusion in viscous liquids is in a state of flux. Formerly considered as a problem in energy-barrier crossing—and hence strictly kinetic in nature—it is now being suggested by molecular-dynamics (MD) studies that liquid diffusion is a process dominated by thermodynamic factors. Nearly three years ago, Sastry *et al.*<sup>4</sup> presented an 'inherent structures' analysis of MD data on a binary Lennard-Jones liquid (BMLJ), showing that two key dynamic features could be related to the static structure through the potential-energy hyper-surface features that the liquid explores with highest probability at different temperatures. These were the onset of 'super-Arrhenius' behaviour, and the location of the mode-coupling theory critical temperature  $T_c$  (obtained by power-law fitting of the diffusivity versus temperature relation). An implication that the landscape excitation profile would provide a measure of liquid fragility that was basically thermodynamic was quantified by Speedy, who derived<sup>5</sup> an expression for thermodynamic fragility in which the excess entropy 'frozen in' at the kinetic glass transition plays an important scaling role.

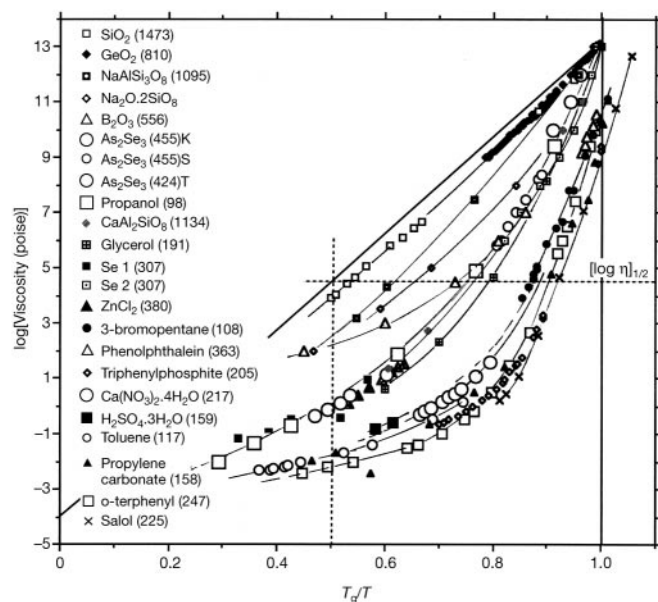
This implication was also followed up by Ito *et al.*<sup>2</sup>, who recast Kauzmann's normalized entropy data in a  $T_g$ -scaled form that has the same appearance as the common kinetic fragility plot. Their plot also assigns a key scaling role to the excess entropy frozen in at  $T_g$ . They observed, qualitatively, that the ordering of liquids was the same in both kinetic and thermodynamic manifestations. This notion was however immediately contested by Ngai *et al.*<sup>3,6</sup>. These authors suggested, from an analysis of the dielectric relaxation times and entropies of nine molecular liquids<sup>3</sup>, that the correlation was unreliable at best and, in the case of polymeric liquids<sup>6</sup>, was not applicable. Here we provide data on a much wider set, which includes all the cases of ref. 3, in order to clarify this problem. Clarification is demanded because, if kinetic fragility is indeed dominated by thermodynamics in the 'normal' case, then a significant simplification of the whole fragility phenomenon



(and also the topography of the potential-energy surface) is at hand.

We start with a proviso about thermodynamic fragility. In the definition of this quantity in ref. 2, one advantage of the kinetic fragility is missing. This is the manner in which the kinetic fragility is made independent of the properties of the crystal phase. By using  $T_g$  as a reference temperature instead of the melting point, the kinetic fragility is made an entirely liquid-state quantity. In the thermodynamic fragility analogue defined by Ito *et al.*<sup>2</sup>, this reference temperature scaling advantage is maintained. Unfortunately, however, in our attempt to select the important part of the entropy for liquid-state behaviour (that is,  $S_{ex}$ , the part in excess of the entropy of the fixed structure, which is usually taken as that of the crystal), we find ourselves still at the mercy of idiosyncrasies of the crystal phase. This is because changes of vibrational entropy on fusion will affect the entropy of fusion, to which  $S_{ex}$  is always referred (except in MD evaluations<sup>7–13</sup>).

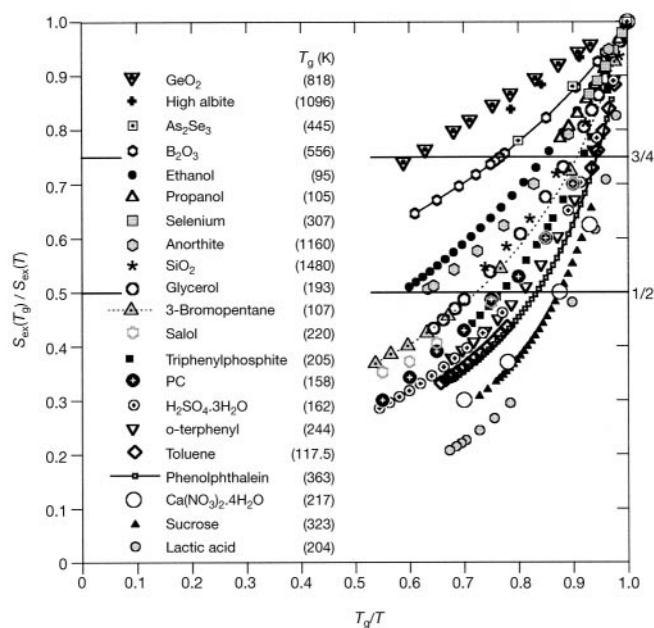
The outstanding case is that of  $\text{SiO}_2$  and certain silicates<sup>14,15</sup> (Fig. 2 legend). A similar problem with chain polymer crystals, in which the force-constant distribution must be even more heterogeneous than in silicate crystals, is perhaps the reason that attempts to correlate Vogel–Fulcher and Kauzmann temperatures, respectively  $T_0$  and  $T_K$ ,



**Figure 1**  $T_g$ -scaled Arrhenius plot for liquids of every class for which thermodynamic data are also available. The scaling parameter used here and in Fig. 2 is the normal calorimetric  $T_g$  at which the enthalpy relaxation time is  $\sim 200$  s (values are shown in parentheses in the key to the figure). The horizontal line is drawn half way between  $10^{13}$  poise (characteristic of the glass transition for non-fragile liquids), and  $10^{-4}$  poise (which is the roughly common high-temperature limiting value). The  $(\log \eta)_{1/2}$  line is used to obtain the  $F_{1/2}$  fragility (kinetic) by the definition  $F_{1/2} = 2T_g/T_{1/2} - 1$ . This is recommended<sup>18</sup> as the most reliable metric of the deviation from Arrhenius behaviour, which is central to the fragility concept. We note that the kinetic fragilities of liquids whose viscosities are less than  $10^{12}$  Pa s at their calorimetric  $T_g$  (due to small shear moduli or viscosity/structure decoupling) are exaggerated by this construction. Fragility plots using relaxation times instead of viscosities are simpler because they all pass through the same point at the reference temperature<sup>3,18</sup>. Relaxation time data, however, are much less generally available than viscosity data, and dielectric relaxation times are not available for ionic or covalent liquids. We have also (L.-M.M. *et al.*, unpublished work) made a modified version of this plot where we take into account the mechanical relaxation times that are available for a smaller set of liquids. (Relaxation times under shear are proportional to viscosities via a Maxwell relation in which the proportionality constant ( $G_\infty$ , the high-frequency shear modulus) has only a weak temperature dependence). The available mechanical relaxation time data will be presented elsewhere.

(see Methods) have never met with much success for polymers. It might also explain why, for polymers, Roland *et al.*<sup>6</sup> found such disarray in the correlation we are discussing here. Until we can obtain the entropy of the glass at  $T_g$  using the ideal-gas reference state, as in MD studies<sup>7–13</sup>, we must expect to find some serious anomalies. We do not think this should prevent us from making an attempt to see the broader picture. We hope that the conclusions about the source of fragility that we report here will justify our tolerance of some cases which, without the proviso just made, might seem to sustain the conclusions of refs 3 and 6. Why not avoid the crystal problem by reverting to the use of the constant-pressure heat capacity ratio  $C_{p,liquid}/C_{p,glass}$ , a purely liquid-state property, as in the past? Speedy's argument<sup>5</sup>, and the Adam–Gibbs equation<sup>16</sup> (the form of which has now been strongly supported by MD simulations<sup>11–13</sup>), show that the scaled entropy quantity is required.

With the foregoing proviso in mind, we will now show that comparison of viscosity data with thermodynamic data on the same liquids broadly confirms the original suggestion<sup>2</sup>. As shown in the



**Figure 2** Plots of scaled excess entropy versus  $T_g$ -scaled inverse temperature. Data are shown for members of every class of glass-forming systems except metals (for which appropriate compound entropy data are unavailable). The excess of liquid entropy over that of the crystal,  $S_{ex}$ , at different temperatures  $T_g/T$  above  $T_g$  is shown scaled by the excess entropy at  $T_g$ ,  $S_{ex}(T_g)$ , so that the figure appears similar to Fig. 1. The lines drawn at the 0.5 and 0.75 marks are used to obtain thermodynamic fragilities  $F_{1/2}$  and  $F_{3/4}$  for comparison with the kinetic quantities. The latter is preferable, as it does not require any data extrapolations (for strong liquids) for its determination. We note the anomalous position of  $\text{SiO}_2$ , which is the strongest liquid in Fig. 1. This anomaly has been pointed out before, by Richet<sup>14,15</sup>, in his consideration of the thermodynamics of geosilicate glass-formers. The anomaly arises because in silica, almost uniquely<sup>14</sup>, the vibrational entropy of the glass is much less than those of the crystalline polymorphs. This has the consequence that the entropy of fusion, and therefore the computed excess entropy of the glass at  $T_g$  (which scales the liquid entropy), is confusingly small. The problem would not arise if the glass entropy could be assessed by either of the all-liquid routes used in MD computer-simulation studies<sup>7–13</sup>. Richet noted that it is not possible to find a systematic relation between the entropy of fusion and the liquid properties in geochemical melts because of such complexities in the crystal thermodynamics. Comparable anomalies, having different origins, attend some other complex silicates. However, Richet showed that, if attention is focused on the increase in heat capacity at  $T_g$  (which, like the viscosity, is a purely liquid state property), then kinetic fragility and thermodynamic fragility, judged by  $C_{p,liquid}/C_{p,glass}$  at  $T_g$ , are systematically related for these mineral liquids—which are mostly 'strong' in character.

Methods section, this is consistent with the earlier finding<sup>17</sup> that the thermodynamic and kinetic ground-state temperatures are mostly equal.

Figures 1 and 2 show plots of the kinetic fragility and thermodynamic fragility of the liquids that are listed in the legends (in order of increasing fragility). We note that the order of the curves in Figs 1 and 2 is the same with few exceptions, though the exceptions may be striking—for example,  $\text{SiO}_2$  in Fig. 2 (see legend).

To make a more quantitative comparison, we may use the intersections of the individual data sets with the mid-range line indicated on each figure. This gives us the  $F_{1/2}$  fragility, the advantages and measurement of which have been discussed elsewhere<sup>18</sup> for the simpler case of relaxation times. For the thermodynamic fragility it is preferable to use the 3/4 line, as this avoids the need for extrapolations in the case of strong liquids.

The thermodynamic and kinetic fragilities, defined in this way, are compared in Fig. 3. While the exceptions are of course still present, most of the substances analysed show thermodynamic fragilities that are well correlated with their kinetic fragilities. The exceptions identified in ref. 3 weigh less heavily among the larger number of simply related cases. We note the wide variety of different liquid types of intermediate fragility that cluster around selenium in Fig. 3.

Figure 3 shows a much closer correlation than we would have expected from the application of the Adam–Gibbs equation<sup>16</sup>, which contains a specifically kinetic parameter (see below). We previously argued<sup>1</sup> that this parameter would exert an independent

control of the fragility but it now seems that the effect is minor. However, before anything can be made of this simplification in both phenomenology and implied configuration-space topography<sup>19,20</sup>, we must face the fact that our correlation is not the one that even a simplified Adam–Gibbs equation would predict. This is because the excess entropy that we have used to construct Fig. 2 is not the quantity that should appear in the Adam–Gibbs equation, even though it is the quantity that has been used in most of its experimental tests. We now analyse the implications of this difference.

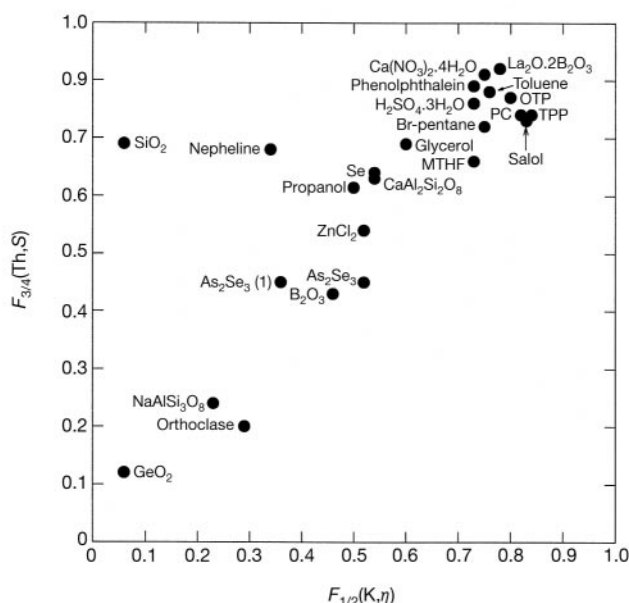
The Adam–Gibbs equation for the relaxation time  $\tau$  of a dense liquid<sup>16</sup>, which has been tested and found to be of valid form in three recent computer simulation studies<sup>7,11–13</sup>, is

$$\tau(\sim TD^{-1}) = \tau_0 \exp(C/TS_c) \quad (1)$$

where  $D$  is the diffusivity (and the  $T$  multiplier takes into account the Einstein relation between diffusivity and mobility),  $\tau_0$  is a constant of order  $10^{-14}$  s,  $C$  is a constant which contains an energy barrier per particle, and  $S_c$  is the configurational entropy.  $S_c$  is that part of the entropy of a pure liquid which is determined by the plethora of possible distinct packing states accessible at the temperature  $T$ . It is related logarithmically to the number of minima on the Goldstein–Stillinger potential-energy hypersurface of the system's configuration space<sup>19,20</sup>. It is accessible by computer simulation studies when correctly performed<sup>7–13</sup>. While it has often been assumed that  $S_c$  can be approximated by the difference in entropy between liquid and crystal, it has been known since Goldstein's analysis<sup>21</sup> of entropy change in glasses between  $T_g$  and 0 K that this is a rather poor approximation in many cases. According to such studies<sup>21,22</sup>, the configurational component of the excess entropy may be as small as 40% of the excess entropy at  $T_g$ , with no pattern obvious for different types of liquids. Sastry's simulations<sup>13</sup> help to explain why. A reason for ignoring these observations has no doubt been the fact that the Adam–Gibbs equation usually tests well (that is, linearizes the relaxation time or diffusivity data near  $T_g$ ) using the easily accessible excess entropy<sup>23–25,17,18</sup>. Figure 3 provides another example of this success. Yet MD simulation studies of both SPC-E water<sup>11</sup> and BMLJ liquid<sup>12,13</sup> find that equation (1) holds as written. In neither case is the excess entropy equal to the configurational entropy. How is it that the Adam–Gibbs equation can be satisfied for both quantities?

The answer to this question must be that, as the temperature rises above  $T_g$ ,  $S_c$  and  $S_{ex}$  change with temperature proportionally (for example,  $S_{ex} = aS_c(T - T_K)$  where  $T_K$  is the Kauzmann vanishing  $S_{ex}$  temperature. When the rate of change is high, the liquid is fragile (Fig. 2). From these two statements we can deduce much about heat capacity and the origin of fragility. An increase in vibrational (or anharmonicity) entropy in excess of the crystal, as the structure changes above  $T_g$ , must promote a corresponding increase in configurational entropy (hence a decrease in  $\tau$ ), at least in the range over which the experimental tests of equation (1)<sup>17,18,23–26</sup> have been made (that is, between  $T_g$  and the melting point at about  $\sim 1.5T_g$ ) (L.M.-M. *et al.*, unpublished results). Here we present the argument graphically, in Fig. 4, using the configuration-space energy landscape concepts as developed in refs 4, 19, 20 and 27. We have recently learned that both S. Sastry (personal communication) and Starr *et al.*<sup>31</sup> have demonstrated in simulations the proportionality between  $S_c$  and  $S_{ex}$  that we have proposed here.

The generation of extra vibrational entropy in the liquid—due to the change in shape of the inherent structure basins (Fig. 4a) that the system 'visits' at higher temperatures—provides an additional drive towards the higher-energy states over that provided by the multiplicity of basins alone. This is dictated by the Gibbs free-energy function  $G = H - TS$ , where  $H$  is enthalpy and  $S$  is entropy, which must be minimized at each temperature for the liquid to be in equilibrium. We note that, because we are discussing the results of Figs 1–3, we are depicting the energy profile for a constant-pressure system, which therefore has one extra dimension—volume—relative



**Figure 3** Correlation of the thermodynamic and kinetic  $F_{1/2}$  fragilities from Figs 1 and 2. Sucrose and lactic acid are excepted for lack of reliable viscosity data. The thermodynamic fragilities of strong liquids, for example,  $\text{SiO}_2$ , are very susceptible to distortion due to vibrational entropy differences of crystal and glass, and would be better assessed using heat capacity jumps at the glass temperature (see Fig. 2 legend). This figure contains additional data ( $\text{ZnCl}_2$ ,  $\text{La}_2\text{O}_3 \cdot 2\text{B}_2\text{O}_3$ , MTHF) not included in Figs 1 and 2 for reason of crowding. The present figure contains all of the examples included in ref. 2, which are here concentrated in the top right-hand corner (glycerol and above). We have added inorganic liquids, for example,  $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ ,  $\text{H}_2\text{SO}_4 \cdot 3\text{H}_2\text{O}$ ,  $\text{La}_2\text{O}_3 \cdot 2\text{B}_2\text{O}_3$ , to this area. In ref. 3, Fig. 3 insert, MTHF is the outstanding exception, 3-bromopentane was misplotted, and only one of the two toluene points has the appropriate excess entropy. In ref. 17, propanol, salol and phenolphthalein were exceptions. On the scale of the present figure these deviations are less prominent. MTHF, methyl tetrahydrofuran; TPP, triphenyl phosphite; OTP, orthoterphenyl. (Th, S) indicates a thermodynamic assessment based on entropy data, and (K,  $\eta$ ) indicates a kinetic assessment based on viscosity data.

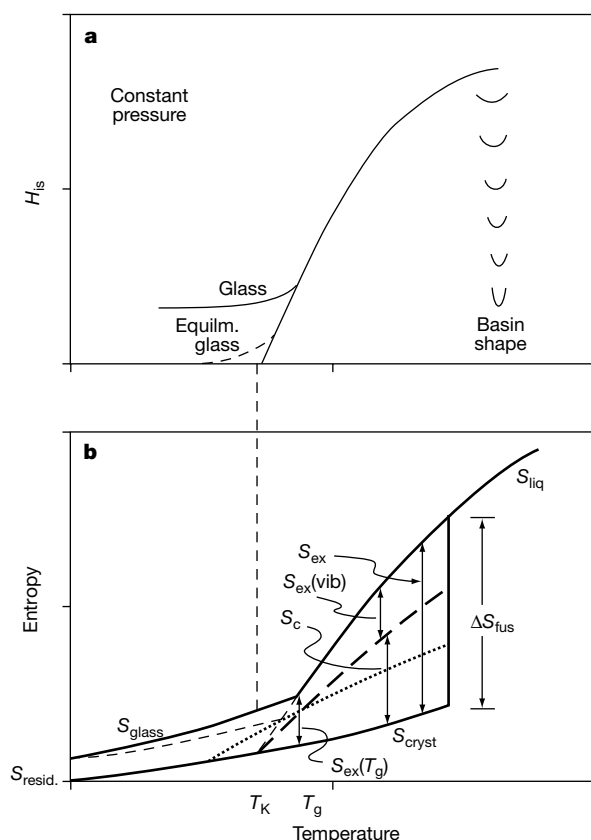
to the usual isochoric landscape (hence we refer to Gibbs, rather than Helmholtz, free energy). The more rapidly the landscape is ascended, the more rapidly the configurational entropy is excited. According to equation (1) this would imply more fragile liquid behaviour, but in fact this depends on the scaling principle used in the comparison and on which of  $S_c$  or  $S_{ex}$  is the most important. The change in basin shape with basin energy is a key ingredient in determining water thermodynamics<sup>28</sup> and diffusivity temperature dependence<sup>11</sup>, and BMLJ aging kinetics<sup>31</sup>. We expect it also to be, along with the other factors  $k_B \ln W$  (where  $W$  is the “number of glasses”<sup>5</sup> and  $k_B$  the Boltzmann constant) and the width of the basin energy distribution identified by Sastry<sup>13</sup>, a key ingredient in determining the liquid fragility.

For studies conducted at constant volume, as is customary in computer simulations<sup>4,8–13,28,31</sup>, a very different scenario is revealed by Sastry’s recent study of BMLJ<sup>13</sup>. There the excess vibrational entropy temperature dependence can play an opposing role. This difference needs to be borne in mind when comparing results of simulations and experiments. It implies for fragile liquids the relation  $C_{p,ex} > C_{c,p} > C_{c,v} > C_{v,ex}$  (where  $C_c$  is the configurational heat capacity).

Figure 4b shows how extrapolations of either total excess entropy, or the configurational component of the excess entropy, should yield the same Kauzmann temperature. It also shows why equation (1) should linearize the data, irrespective of whether excess entropy or configurational entropy is used, and also why—in correlations of thermodynamic fragility with the kinetic quantity—excess entropy should serve as well as configurational entropy. Experimental verification of the qualitative features of Fig. 4b is available in the neutron-scattering study of selenium by Phillips *et al.*<sup>29</sup> (see their Fig. 4). The origin of the excess vibrational entropy is there seen to lie in changes of the density of states (DOS) in the vicinity of the boson peak, as may also be deduced from refs 21 and 22. In  $\text{SiO}_2$  (the strong extreme), the opposite changes in the DOS above  $T_g$  are found<sup>30</sup>. □

## Methods

We support our claim that kinetic, and entropy-based thermodynamic, fragilities are closely related, using three different approaches as detailed below. These involve both experimental and MD computer-simulation data sources. Only the third, experiment-based, approach is detailed here.



**Figure 4** Graphical explanation of how high liquid fragility originates in the change of energy landscape ‘basin shape’ (hence of the vibrational density of states) during thermal excitation to high potential energies. **a**, Depiction of the inherent structure enthalpy  $H_{is}$  profile for a one-component system during heating from 0 K to above the melting point at constant pressure. Changes in shape of the basins visited with greatest probability<sup>4</sup> at each temperature and volume are depicted to the right of the excitation profile. These become progressively shallower (density of states slanted to lower frequencies) as temperature increases and the system expands. **b**, Entropy versus temperature for the system depicted in **a**, contrasting the monotonic increase of the system entropy,  $S_{cryst}$ , when in the crystal basin, with the behaviour of amorphous phases, both strong and fragile. The glass is depicted as having excess entropy  $S_{resid}$  at 0 K but a larger  $S_{ex}$  at  $T_g$ . This is due to the vibrational density of states of the glass being richer in low-frequency modes than that of the crystal. Above  $T_g$  the vibrational excess entropy  $S_{ex}(vib)$  increases more rapidly than below  $T_g$  due to the system having access to basins of different

shape as depicted in **a**. This access to additional vibrational entropy provides an additional  $TS$  drive which must be balanced by more rapid increase in enthalpy  $H$ , than in the case where all basins have the same shape (dotted line). The division of the excess entropy into its configurational and vibrational components is made by the heavy dashed line. In the case of constant basin shape (strong liquid), the drive to higher enthalpies comes only from the configurational entropy, that is,  $TS_c$ . The change in  $S_{ex}$ , which is also  $S_c$  in this case, is depicted as a dotted line. We note that, for variable basin shape, the increased rate of excitation of vibrational entropy has increased the rate of generation of configurational entropy, as well as that of the excess entropy.  $T_g/T_K$  must therefore be smaller for this case, hence the fragility larger. Compare with Fig. 4 of ref. 29. The increase in entropy at the melting point  $\Delta S_{fus}$ , to give the liquid entropy  $S_{liq}$ , is seen to contain both configurational and vibration-related components in the case of fragile liquids.



(1) The correlation of kinetic and entropy-based thermodynamic fragilities is implied by the repeated analyses of the calorimetric and kinetic behaviour of liquids (not polymers) of different classes that extract 'ground state temperatures' from the data and find them to have very similar values. We refer to  $T_K$  (where the supercooling liquid entropy extrapolates to the crystal entropy) and  $T_0$  (where the viscosity extrapolates to infinity by the Vogel–Fulcher–Tammann equation). Ref. 17 describes some 30 cases where  $T_0 = T_K$  within  $\pm 3\%$  despite several pronounced exceptions. As the ratio  $T_g/T_0$  is often taken as a measure of fragility<sup>17</sup>, it is clear that  $T_K \approx T_0$  implies that kinetic fragility and thermodynamic fragility  $T_g/T_K$  are the same, with some pronounced exceptions. This means that either the physically constrained analysis leading to  $T_0 \approx T_K$  is flawed, or that the analysis of Ngai<sup>3</sup> and co-workers, being limited to dielectric relaxation of molecular liquids, and containing some misplots (see Fig. 3 legend), has distorted the overall picture.

(2)  $T_0 \approx T_K$  is also a recent result of MD computer-simulation studies on well defined systems<sup>11–13</sup>. Although the number of systems studied is small, the results are important because the identity of  $T_0$  and  $T_K$  is obtained using all-amorphous phase calculations; that is, there is no reliance on fusion entropy data.

(3) Both of the above arguments are weakened by the fact that they depend on extrapolations beyond the range of experimental or simulation data. To avoid this, and at the same time to obtain a more complete picture than given in refs 2 and 3, we compare wide-ranging data on some 25 liquids from the fields of geochemistry (liquid silicates), covalent semiconductors (liquid chalcogenides), molecular liquids, and molten ionic hydrates, salts and oxides, and including all the systems analysed by Ngai and Yamamuro<sup>3</sup>. Unlike the cases in ref. 3, the present data set covers the whole known fragility range. The quantities studied are the shear viscosities and the excess of the liquid entropy over the crystal entropy measured at the same temperature, and uncorrected in any manner. We note that  $S_{ex}$  in ref. 22 is not the same quantity as in the present work. It requires quantitative knowledge of the residual entropy at 0 K for its assessment. Also, in ref. 26 an attempt was made to separate out vibrational contributions to the entropy of the liquid above  $T_g$  in order to obtain  $S_c$  (for use in the Adam–Gibbs equation (equation (1)) by using a function fitted to the heat capacity of the glass. This yields a quantity which is neither  $S_c$  nor  $S_{ex}$  because the procedure does not consider the introduction of new low frequencies above  $T_g$ , which is central to our conclusions.

The quality of data used in this study is variable, ranging from very high, in the case of adiabatic calorimetry data on molecular liquids (data cited in ref. 3) and some high temperature substances ( $B_2O_3$ ), to moderate, in the case of high-temperature differential scanning calorimeter (DSC) and drop calorimetry data on minerals. Reference sources are available from the authors. Both data sets are presented (Figs 1 and 2) as functions of inverse temperature scaled by the calorimetric onset  $T_g$  for the substance, as measured during upscan at  $10\text{ K min}^{-1}$  (after previous continuous cooling at the same rate). This corresponds to scaling by the temperature at which the enthalpy relaxation time is approximately 200 s (refs 17, 18).

Received 10 November 2000; accepted 29 January 2001.

- Angell, C. A. Relaxation in liquids, polymers and plastic crystals - strong/fragile patterns and problems. *J. Non-Cryst. Solids* **131**–133, 13–31 (1991).
- Ito, K., Moynihan, C. T. & Angell, C. A. Thermodynamic determination of fragility in liquids and a fragile-to-strong liquid transition in water. *Nature* **398**, 492–495 (1999).
- Ngai, K. L. & Yamamuro, O. Thermodynamic fragility and kinetic fragility in supercooled liquids: a missing link. *J. Chem. Phys.* **111**, 10403–10406 (1999).
- Sastry, S., Debenedetti, P. G. & Stillinger, F. H. Signatures of distinct dynamical regimes in the energy landscape of a glassforming liquid. *Nature* **393**, 554–557 (1998).
- Speedy, R. J. Relations between a liquid and its glasses. *J. Phys. Chem. B* **103**, 4060–4065 (1999).
- Roland, C. M., Santangelo, P. G. & Ngai, K. L. The application of the energy landscape model to polymers. *J. Chem. Phys.* **111**, 5593–5598 (1999).
- Speedy, R. J. The hard sphere glass transition. *Mol. Phys.* **95**, 169–178 (1998).
- Sciortino, F., Kob, W. & Tartaglia, P. Inherent structure entropy of supercooled liquids. *Phys. Rev. Lett.* **83**, 3214–3217 (1999).
- Buechner, S. & Heuer, A. The potential energy landscape of a model glassformer: thermodynamics, anharmonicities, and finite size effects. *Phys. Rev. E* **60**, 6507–6518 (1999).
- Coluzzi, B., Verrocchio, P., Mezard, M. & Parisi, G. Lennard-Jones binary mixture: a thermodynamical approach to glass transition. *J. Chem. Phys.* **112**, 2933–2944 (2000).
- Scala, A., Starr, F., La Nave, E., Sciortino, F. & Stanley, H. E. Configurational entropy and diffusion of supercooled water. *Nature* **406**, 166–169 (2000).
- Sastry, S. Liquid limits: the glass transition and liquid-gas spinodal boundaries of metastable liquids. *Phys. Rev. Lett.* **85**, 590–593 (2000).
- Sastry, S. The relationship between fragility, configurational entropy and the potential energy landscape of glassforming liquids. *Nature* **409**, 164–167 (2001).
- Richet, P. Viscosity and configurational entropy of silicate melts. *Geochim. Cosmochim. Acta* **48**, 471–483 (1984).
- Richet, P. & Bottinga, Y. Glass transitions and thermodynamic properties of amorphous  $SiO_2$ ,  $NaAlSi_3O_8$  and  $KAlSi_3O_8$ . *Geochim. Cosmochim. Acta* **48**, 453–470 (1984).
- Adam, G. & Gibbs, J. H. On the temperature dependence of cooperative relaxation properties in glassforming liquids. *J. Chem. Phys.* **43**, 139–146 (1965).
- Angell, C. A. Entropy and fragility in supercooled liquids. *J. Res. NIST* **102**, 171–185 (1997).
- Richet, P. & Angell, C. A. Dynamics of glassforming liquids. IV: On the link between molecular dynamics and configurational entropy. *J. Chem. Phys.* **108**, 9016–9026 (1998).
- Goldstein, M. Viscous liquids and the glass transition: a potential energy barrier picture. *J. Chem. Phys.* **51**, 3728–3729 (1969).
- Stillinger, F. H. & Weber, T. A. Packing structures and transitions in liquids and solids. *Science* **225**, 983–989 (1984).
- Goldstein, M. Viscous liquids and the glass transition: sources of the excess heat capacity. *J. Chem. Phys.* **64**, 4767–4773 (1976).
- Johari, G. P. Contributions to the entropy of a glass and liquid, and the dielectric relaxation time. *J. Chem. Phys.* **112**, 7518–7523 (2000).

- Greet, R. J. & Turnbull, D. Test of Adam–Gibbs liquid viscosity model with *o*-terphenyl specific-heat data. *J. Chem. Phys.* **47**, 2185–2190 (1967).
- Magill, J. H. Physical properties of aromatic hydrocarbons. III. A test of the Adam–Gibbs relaxation model for glass formers based on the heat-capacity data of 1,3,5-tri- $\alpha$ -naphthylbenzene. *J. Chem. Phys.* **47**, 2802–2807 (1967).
- Takahara, S., Yamamuro, O. & Suga, H. Heat capacities and glass transitions of 1-propanol and 3-methyl pentane: new evidence for the entropy theory. *J. Non-Cryst. Solids* **171**, 259–270 (1994).
- Takahara, S., Yamamuro, O. & Matsuo, T. Calorimetric study of 3-bromopentane—correlation between structural relaxation time and configurational entropy. *J. Phys. Chem.* **99**, 9580–9592 (1995).
- Angell, C. A. Liquid landscapes. *Nature* **393**, 521–522 (1998).
- Starr, F. W. *et al.* Thermodynamic and structural aspects of the potential energy surface of simulated water. *Phys. Rev. E* (in the press); also preprint arXiv:cond-mat/0007487 on (xxx.lanl.gov) (2000).
- Phillips, W. A., Buchenau, U., Nücker, N., Dianoux, A. J. & Petry, W. Dynamics of glassy and liquid selenium. *Phys. Rev. Lett.* **63**, 2381–2384 (1989).
- Wischniewski, A., Buchenau, U., Dianoux, A. J., Kamitakahara, W. A. & Zarestky, J. L. Neutron scattering analysis of low-frequency modes in silica. *Phil. Mag. B* **77**, 579–589 (1998).
- Kob, W., Sciortino, F. & Tartaglia, P. Aging as dynamics in configuration space. *Europhys. Lett.* **49**, 590–596 (2000).

## Acknowledgements

We thank S. Sastry, R. Speedy and F. Sciortino for comments and criticisms. This work was supported by the NSF, Division of Materials Research, Solid State Chemistry program.

Correspondence and requests for materials should be addressed to C.A.A. (e-mail: caa@asu.edu).

# Intermittent dislocation flow in viscoplastic deformation

M.-Carmen Miguel\*†, Alessandro Vespignani\*, Stefano Zapperi‡, Jérôme Weiss§ & Jean-Robert Grasso||

\* The Abdus Salam International Centre for Theoretical Physics, PO Box 586, 34100 Trieste, Italy

† Departament de Física Fonamental, Facultat de Física, Universitat de Barcelona, Av. Diagonal 647, 08028 Barcelona, Spain

‡ INFN, Università “La Sapienza”, P. le A. Moro 2, 00185 Roma, Italy

§ LGGE-CNRS, 54 rue Molière, BP 96, 38402 St Martin d'Hères Cedex, France

|| LGIT, BP 53X, 38041 Grenoble Cedex 9, France

The viscoplastic deformation (creep) of crystalline materials under constant stress involves the motion of a large number of interacting dislocations<sup>1</sup>. Analytical methods and sophisticated ‘dislocation dynamics’ simulations have proved very effective in the study of dislocation patterning, and have led to macroscopic constitutive laws of plastic deformation<sup>2–9</sup>. Yet, a statistical analysis of the dynamics of an assembly of interacting dislocations has not hitherto been performed. Here we report acoustic emission measurements on stressed ice single crystals, the results of which indicate that dislocations move in a scale-free intermittent fashion. This result is confirmed by numerical simulations of a model of interacting dislocations that successfully reproduces the main features of the experiment. We find that dislocations generate a slowly evolving configuration landscape which coexists with rapid collective rearrangements. These rearrangements involve a comparatively small fraction of the dislocations and lead to an intermittent behaviour of the net plastic response. This basic dynamical picture appears to be a generic feature in the deformation of many other materials<sup>10–12</sup>. Moreover, it should provide a framework for discussing fundamental aspects of plasticity that goes beyond standard mean-field approaches that see plastic deformation as a smooth laminar flow.

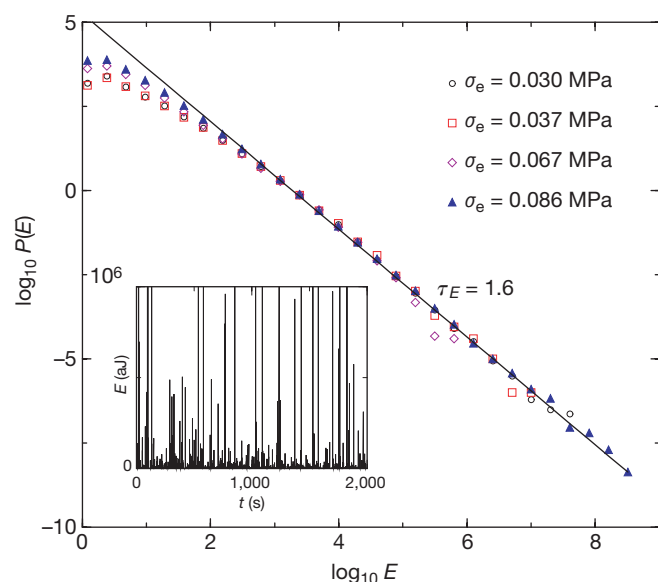
Whenever dislocation glide is the dominant plastic deformation mechanism in a crystalline material, we observe a constant strain-rate regime usually described by Orowan’s relation  $\dot{\gamma} \approx \rho_m b v$ . Here, the plastic strain-rate of the material  $\dot{\gamma}$  is simply related to average quantities such as  $\rho_m$ , the density of mobile dislocations, and  $v$ , their

average velocity along the slip direction (parallel to the Burgers vector  $b$ )<sup>1</sup>. Transmission electron micrographs of plastically deformed materials display, on the other hand, complex features such as cellular structures and fractal patterns<sup>2,3</sup>, which are the fingerprint of a complex multiscale dynamics not appropriately accounted for by the mean-field character of Orowan's relation. In addition, rapid slip events<sup>10</sup> have been observed in the plastic deformation of various metals and alloys<sup>11,12</sup>, and in the Portevin–LeChatelier effect<sup>13</sup>. We believe that formulating plastic deformation as a nonequilibrium statistical mechanics problem<sup>14</sup> requires a substantial understanding of basic collective dislocation dynamics.

Experimentally, the complex character of collective dislocation dynamics can be revealed by acoustic emission measurements. The acoustic waves recorded in a piezoelectric transducer disclose the pulse-like changes of the local displacements in the material during plastic deformation, whereas a smooth plastic flow would not be detected<sup>15</sup>. Thus, this method is particularly useful for inspecting possible fluctuations in the dislocation velocities and densities.

Ice single crystals can be used as a model material to study glide dislocation dynamics by acoustic emission (refs 16, 17) for the following reasons: (1) Transparency allows direct verification that acoustic emission activity is not related to microcracking. (2) Within the range of temperature and stress corresponding to our experimental conditions, diffusional creep is not a significant mechanism of inelastic deformation<sup>18</sup> which, in hexagonal ice single crystals, occurs essentially by dislocation glide on the basal planes along a preferred slip direction. (3) An excellent coupling between sample and transducer can be obtained by fusion/freezing.

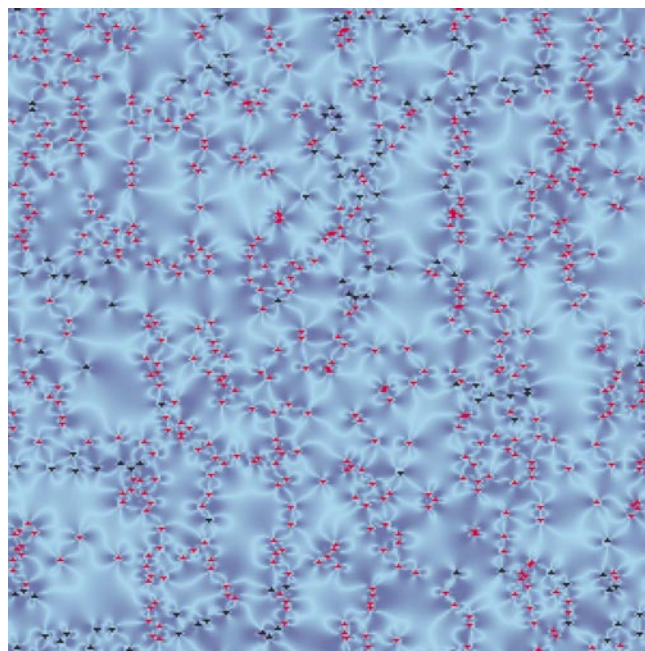
Uniaxial compression creep experiments are performed on artificial ice single crystals, employing several steps of constant applied stress. We observe an intense acoustic activity, exhibiting a strong intermittent character (see Fig. 1 inset). We measure the energy associated with each acoustic burst and analyse its statistical proper-



**Figure 1** Statistical properties of the acoustic energy bursts recorded in ice single crystals under constant stress. The main figure shows the distribution of energy bursts for the different loading steps. The cut-offs depend slightly on the applied stress, but the power-law exponent remains the same. The fit yields an exponent  $\tau_E = 1.60 \pm 0.05$ . Inset, a typical recorded acoustic signal. The creep experiment is performed at  $T = -10^\circ\text{C}$  under different constant compression stresses ranging from  $\sigma = 0.58\text{ MPa}$  to  $\sigma = 1.64\text{ MPa}$ . The stress is applied at an angle with respect to the  $c$  axis, giving rise to a small resolved shear stress acting across the glide plane ( $\sigma_s = 0.030\text{ MPa}$ – $0.086\text{ MPa}$ ). The frequency bandwidth of the transducer was 10–100 kHz. The amplitude range between the minimum and the maximum recordable thresholds is 70 dB, that is, a range of seven orders of magnitude in energy.

ties. In Fig. 1 we show that the probability distribution of energy burst intensities exhibits a power-law behaviour spanning several decades. This fact is an indication that a large number of dislocations move cooperatively in an intermittent fashion. A similar behaviour has been observed in the Portevin–LeChatelier effect<sup>13</sup>, a plastic instability where the intermittent flow is ruled by the interaction of dislocations and diffusing solute atoms. The intermittency observed in our case is different as we only have interacting dislocations subject to an external stress, without any other alien element interfering with their dynamics.

Multiscale properties and pattern formation are ubiquitous in plastic materials and we expect that the large dynamical fluctuations observed in ice compression experiments are also a prevalent feature of plastic deformation micromechanics. This general picture implies that the experimental phenomenology should be reproducible in simple numerical simulations of discrete dislocation dynamics that preserve the relevant characteristics of the system under study. As in previous dislocation dynamics models<sup>4–7</sup>, we consider a two-dimensional cross-section of the crystal (that is, the  $xy$  plane), and randomly place  $N$  straight-edge dislocations gliding along a single slip direction parallel to their respective Burgers vectors  $\mathbf{b}$  (that is, the  $x$  direction). This implies that we have point-like dislocations moving along fixed lines parallel to the  $x$  axis. This simplification effectively describes materials like ice crystals that, owing to their strong plastic anisotropy, deform by glide on a single



**Figure 2** Snapshot of the total stress field and arrangement of dislocations in a numerical simulation with  $v_s = 0.025$ . We observe metastable structure formation (dipoles and walls) and the associated stress field which goes from light blue (low values) to dark blue (high values). The complex low-stress pathways joining dislocations are the result of the anisotropic elastic interactions. Dislocations moving at low velocities (lower than  $v_s$ ) are depicted in red, while those moving at higher velocities are depicted in black. Most dislocations in walls are moving slowly, whereas isolated ones tend to move at higher speed. In the simulations, an initial number of  $N_0 = 1,500$  dislocations are randomly distributed on a square cell of size  $L = 300$ . Their Burgers vectors are randomly chosen to be  $+b$  or  $-b$  with equal probability. In the absence of external stress, we first let the system relax until it reaches a metastable arrangement. The number of remaining dislocations is then  $N \approx 700$ . At this point, we apply a constant shear stress and study their evolution. To avoid the discontinuities arising from truncating long-range elastic interactions in equation (1), we resort to the Ewald summation method. We have thus exactly taken into account the interaction of a dislocation with all the infinite periodic images of another dislocation in the same cell.

slip system. An edge dislocation with Burgers vector  $\mathbf{b}$  located at the origin gives rise to a shear stress  $\sigma_s$  at a point  $\mathbf{r} = (x, y)$  of the form

$$\sigma_s(\mathbf{r}) = b\mu \frac{x(x^2 - y^2)}{2\pi(1 - \nu)(x^2 + y^2)^2} \quad (1)$$

where  $\mu$  is the shear modulus and  $\nu$  is the Poisson ratio<sup>1</sup>. This long-range stress is responsible for mutual interactions among dislocations which are important in all dislocation dynamics models<sup>4–7</sup>. Under a constant external stress  $\sigma_e$ , dislocation  $i$  performs an overdamped motion along the  $x$  direction described by the following equation

$$\eta \frac{dx_i}{dt} = b_i \left( \sum_{m \neq i} \sigma_s(\mathbf{r}_m - \mathbf{r}_i) - \sigma_e \right) \quad (2)$$

where  $\eta$  is the effective friction and  $b_i$  is the Burgers vector. Throughout the simulations, we will be dealing with dimensionless quantities after setting the distance scale  $b = 1$ , and the timescale  $t_0 \equiv \eta b / (\mu / 2\pi(1 - \nu)) = 1$ .

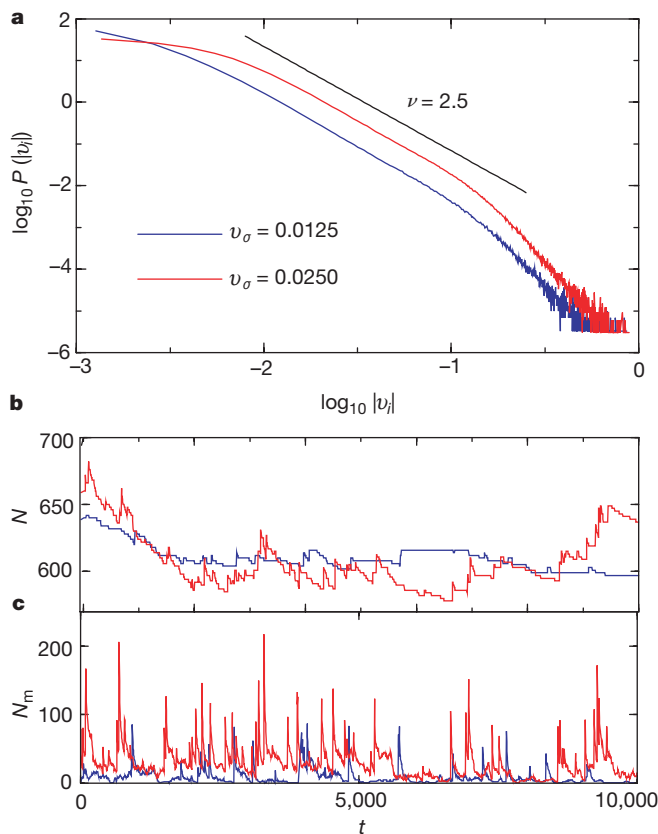
Other essential ingredients commonly introduced in most dislocation dynamics models<sup>4–7</sup> are the mechanisms for dislocation (1) annihilation and (2) multiplication. (1) When the distance between two dislocations is of the order of a few Burgers vectors, the high stress and strain conditions close to the dislocation core invalidate the results obtained from a linear elasticity theory (equation (1)). In these instances, phenomenological nonlinear reactions describe more accurately the real behaviour of dislocations in a crystal. In our model, we annihilate two dislocations with opposite Burgers vectors when the distance between them is shorter than  $2b$ . (2) The

activation of Frank–Read sources<sup>1</sup> is accepted as the most relevant dislocation multiplication mechanism under creep deformation. The activation of Frank–Read sources has been observed in ice crystals, along with more specific multiplication processes<sup>19</sup>. Because an accurate multiplication mechanism cannot be simulated in a two-dimensional point-dislocation model, we have implemented various phenomenological procedures. A simple way of taking into account the new dislocation loops generated at different sources within the crystal is the random introduction of opposite-sign dislocation pairs in the cross-section under consideration. The rate and frequency of this creation process depend solely on the external stress  $\sigma_e$ . We have checked that other multiplication rules in which the creation rate depends on the local stress yield similar results.

In the initial stage of the simulations, the system relaxes slowly and the average strain-rate decreases until it reaches a constant value which depends on the applied stress, that is the stationary creep regime. By monitoring the activity in a single run, we observe that most dislocations are arranged into metastable structures (such as walls and cells) moving at a very slow rate (see Fig. 2 and Supplementary Information). A smaller fraction of dislocations, however, move intermittently at much higher velocities, giving rise to sudden increases of the plastic strain. In Fig. 3a, we plot the probability density of finding an individual dislocation moving with a velocity  $|v_i|$ . Already at the level of individual dislocations, the velocity distribution  $P(|v_i|)$  is quite broad and exhibits power-law behaviour for velocity values larger than the external stress-induced velocity  $v_o = b\sigma_e/\eta$  (we have considered two values of  $v_o$ : 0.0125 and 0.025).

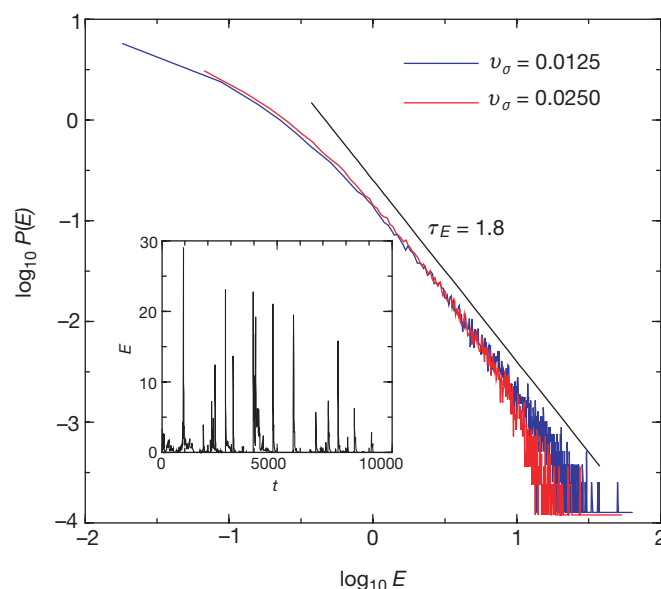
In the presence of a large number of dislocations, the acoustic signal that is detected experimentally is due to the superposition of the waves emitted by each moving dislocation<sup>15</sup>. In particular, it has been shown that a single dislocation performing a sudden movement with velocity  $v_0$  for a short time  $\tau$  gives rise to an acoustic wave whose amplitude is proportional to  $v_0$  (ref. 15). The high-amplitude pulses detected in the experiment cannot be ascribed to uncorrelated emissions from each individual dislocation, but rather to the cooperative motion of several dislocations occurring, for example, after the activation of a Frank–Read source, or if a dislocation cluster breaks apart. To gain further insight into this behaviour, in Fig. 3b and c, we show the evolution of the total number of dislocations as well as that of the fast-moving dislocations, that is, dislocations moving faster than if they were moving only under the action of the external stress,  $v_i > v_o$  (other threshold values yield equivalent results). After the injection of a few new dislocations or the annihilation of a dislocation pair, several other dislocations start to move and rearrange, not necessarily in the close vicinity of the triggering event. To quantify this effect, we measure the collective velocity  $V = \Sigma' |v_i|$  of the fast-moving dislocations and define the acoustic energy as  $E = V^2$  (ref. 17). In the inset of Fig. 4, we see that the signal  $E(t)$  consists of a succession of intermittent and pronounced bursts, each one signalling the onset of collective dislocation rearrangements. The slow and continuum motion of dislocation structures ( $v_i < v_o$ ) is not considered as it only sets up a background noise signal which cannot be detected experimentally. In Fig. 4 we show that the distribution of energy bursts, obtained by sampling the signal over different times and realizations, decays very slowly, spanning various decades. For intermediate values, the distribution shows an algebraic decay with an exponent  $\tau_E = 1.8 \pm 0.2$ , in reasonable agreement with experiments (see Fig. 1). The maximum number of dislocations we can handle in our simulations severely restricts the maximum value of the signal in the system. Consequently, the extension of the power-law regime grows with the number of dislocations, but is eventually bounded by the different nature of the extremes statistics.

The broadly distributed intermittency resulting from the statistical analysis can be interpreted as a nonequilibrium transport



**Figure 3** Statistical properties of dislocation velocities and density obtained in numerical simulations. **a**, Probability density of the individual dislocation velocities  $|v_i|$  in the stationary state. The black line is a least-squares fit of the intermediate data points. We obtain an exponent  $\nu = 2.5$  for the scaling of the intermediate velocities. **b**, Time evolution of the total number of dislocations in the cell. **c**, Fraction of fast-moving dislocations for a given run of the simulations.





**Figure 4** Statistical properties of the energy bursts obtained in numerical simulations. In the model, we define the acoustic energy  $E = V^2$  from the collective velocity  $V = \Sigma_i |v_i|$  of the fast-moving dislocations. The time evolution of  $E$  for a given simulation run is plotted in the inset. The energy distribution averaged over time and over 100 realizations

with different initial conditions is plotted in the main graph. The black line represents the best fit of the intermediate data points. We obtain an exponent  $\tau_E = 1.8 \pm 0.2$  for both values of the external stress.

phenomenon with scaling properties. Scaling behaviour is usually associated with the proximity of a critical point and is characterized by a high degree of universality. Critical exponents only depend on a few fundamental parameters such as the effective dimensionality and the basic symmetries of the system. Specifically, the reasonable quantitative agreement between our model and the experimental data is due to the strong plastic anisotropy present in ice single crystals, which can thus be well described by a two-dimensional model. Although the quantitative results we obtain should be restricted to the case of single slip systems, the general features of the observed intermittent flow regime appear to be generic in plastic deformation<sup>11,12</sup>. The particular universality class will depend on the effective dimensionality of each specific material as well as the particular deformation mechanism, and one can still expect to obtain the right critical exponents from simplified models if the relevant symmetries are correctly taken into account.

The close-to-criticality nonequilibrium framework suggests possible analogies with elastic manifolds driven in random media such as fluid flow in porous media, vortices in superconductors, and charge density waves<sup>20</sup>. In all of these systems, a critical value of the driving force separates a static regime from a moving one, and scaling is observed only close to this point. The intermittent behaviour close to criticality is in these cases associated with static random heterogeneities which exert space-dependent pinning forces on the moving objects. On the contrary, the present dislocation dynamics model, as well as the ice experiment, does not contain any quenched source of pinning forces. Dislocations themselves, through the various structures such as dipoles and walls, generate a pinning force landscape in which the dynamics is virtually frozen, that is, a slow dynamics state. Creation and annihilation mechanisms, often triggered by the presence of unsettled dislocations, allow the system to jump between slow dynamics states through bursts of activity. This behaviour is reminiscent of driven-dissipative self-organizing systems that achieve criticality in the limit of very slow driving<sup>1</sup>.

This emerging picture poses many basic theoretical questions. Is there a critical stress value below which the system decays into a slow dynamics state? Can the large dynamical fluctuations be associated

with a diverging response function and thus be related to a critical point? Can we derive scaling laws relating the various observed exponents, as suggested by the theory of critical phenomena? The answers to all these questions pave the way to the nonequilibrium statistical theory of dislocation motion that is needed for a deeper understanding of the micromechanics of plastic deformation. □

Received 26 October 2000; accepted 26 January 2001.

- Hirth, J. P. & Lothe, J. *Theory of Dislocations* (Krieger, Malabar, Florida, 1992).
- Hähner, P., Bay, K. & Zaiser, M. Fractal dislocation patterning during plastic deformation. *Phys. Rev. Lett.* **81**, 2470–2473 (1998).
- Zaiser, M., Bay, K. & Hähner, P. Fractal analysis of deformation-induced dislocation patterns. *Acta Mater.* **47**, 2463–2476 (1999).
- Lepinoux, J. & Kubin, L. P. The dynamic organization of dislocation structures: A simulation. *Scr. Metall.* **21**, 833–838 (1987).
- Amodeo, R. J. & Ghoniem, N. M. Dislocation dynamics. I. A proposed methodology for deformation micromechanics. *Phys. Rev. B* **41**, 6958–6967 (1990).
- Groma, I. & Pawley, G. S. Computer simulation of plastic behaviour of single crystals. *Phil. Mag.* **A 67**, 1459–1470 (1993).
- Fournet, R. & Salazar, J. M. Formation of dislocation patterns: Computer simulations. *Phys. Rev. B* **53**, 6283–6290 (1996).
- Gil Sevillano, J., Bouchaud, E. & Kubin, L. P. The fractal nature of gliding dislocation lines. *Scr. Metall. Mater.* **25**, 355–360 (1991).
- Thomson, R. & Levine, L. Theory of strain percolation in metals. *Phys. Rev. Lett.* **81**, 3884–3887 (1998).
- Neuhäuser, H. in *Dislocations in Solids* (ed. Nabarro, F. R. N.) 319–440 (North-Holland, Amsterdam, 1983).
- Becker, R. & Orowan, E. Über sprunghafte Dehnung von Zinkkristallen. *Z. Phys.* **79**, 566–572 (1932).
- Bengus, V. Z., Komnik, S. N. & Shitilman, O. B. Dislocation multiplication as a controlling factor of work-hardening. *Phys. Stat. Sol.* **14**, 215–222 (1966).
- Ananthakrishna, G. et al. Crossover from chaotic to self-organized critical dynamics in jerky flow of single crystals. *Phys. Rev. E* **60**, 5455–5462 (1999).
- Hähner, P. On the foundations of stochastic dislocation dynamics. *Appl. Phys. A* **62**, 473–481 (1996).
- Roubey, D., Fleischman, P. & Duvergier, C. Un modèle de source d'émission acoustique pour l'analyse de l'émission continue et de l'émission par saccades. I. Analyse théorique. *Phil. Mag.* **A 47**, 671–687 (1983).
- Weiss, J. & Grasso, J. R. Acoustic emission in single crystals of ice. *J. Phys. Chem. B* **101**, 6113–6117 (1997).
- Weiss, J., Lahaie, F. & Grasso, J. R. Statistical analysis of dislocation dynamics during viscoplastic deformation from acoustic emission. *J. Geophys. Res.* **105**, 433–442 (2000).
- Duval, P., Ashby, M. F. & Andermann, I. Rate-controlling processes in the creep of polycrystalline ice. *J. Phys. Chem.* **87**, 4066–4074 (1983).
- Petrenko, V. F. & Whitworth, R. W. Structure of ordinary ice  $I_h$ . Part II: Defects in ice. Vol. 2: Dislocations and plane defects. (US Army Cold Regions Research and Engineering Laboratory Special Report 94-12, Hanover, New Hampshire, 1994).

20. Kardar, M. Nonequilibrium dynamics of interfaces and lines. *Phys. Rep.* **301**, 85–112 (1998).  
 21. Jensen, H. J. *Self-Organized Criticality* (Cambridge Univ. Press, Cambridge, 1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

#### Acknowledgements

We thank R. Pastor-Satorras, M. Rubí, A. Scala and M. Zaiser for useful discussions, and O. Brisaud, and F. Dominé for help in the preparation of single crystals. We acknowledge partial support from the European Network contract on "Fractal Structures and Self-organization". J.W. is supported by the "Action thématique innovante" of Institut National des Sciences de l'Université-CNRS. Acoustic emission monitoring devices were financed by Université Joseph Fourier.

Correspondence and requests for materials should be addressed to M.-C.M. (e-mail: carmen@ffn.ub.es).

## Varied pore organization in mesostructured semiconductors based on the $[\text{SnSe}_4]^{4-}$ anion

Pantelis N. Trikalitis\*, K. Kasthuri Rangan\*, Thomas Bakas† & Mercouri G. Kanatzidis\*

\* Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA

† Department of Physics, University of Ioannina, Ioannina, 45110, Greece

Open framework metal chalcogenide solids, with pore sizes in the nano- and mesoscale, are of potentially broad technological and fundamental interest in research areas ranging from optoelectronics to the physics of quantum confinement<sup>1,2</sup>. Although there have been significant advances in the design and synthesis of mesostructured silicas<sup>3,4</sup>, the construction of their non-oxidic analogues still remains a challenge. Here we describe a synthetic strategy that allows the preparation of a large class of mesoporous materials based on supramolecular assembly of tetrahedral Zintl anions  $[\text{SnSe}_4]^{4-}$  with transition metals in the presence of cetylpyridinium (CP) surfactant molecules. These mesostructured semiconducting selenide materials are of the general formulae  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$  (where  $1.0 < x < 1.3$ ; M = Mn, Fe, Co, Zn, Cd, Hg). The resulting materials are open framework chalcogenides and form mesophases with uniform pore size (with spacings between 35 and 40 Å). The pore arrangement depends on the synthetic conditions and metal used, and include disordered wormhole, hexagonal and even cubic phases. All compounds are medium bandgap semiconductors (varying between 1.4 and 2.5 eV). We expect that such semiconducting porous networks

could be used for optoelectronic, photosynthetic and photocatalytic applications.

The discovery of a general synthetic pathway to ordered mesoporous silicates by the Mobil group<sup>3,4</sup> led to new hybrid mesoporous solids and has captured the imagination of many materials scientists. During the past decade, assembling inorganic molecular species in the presence of long-chain organic liquid-crystal templates has been used extensively to construct mesoporous oxides. In general, the synthetic approaches involve the organized polymerization of the basic tetrahedral  $[\text{SiO}_4]^{4-}$  anion in the presence of various surfactants<sup>5–7</sup>. Non-oxide porous solids, however, are also important<sup>8,9</sup>, because they may impart electronic properties to the framework, and thus make mesoporous solids into nano-electronic materials<sup>7,10</sup>. Nevertheless, rational synthetic routes to such systems remain a challenge. A promising pathway for the construction of non-oxide mesoporous solids such as sulphides, selenides or tellurides is the supramolecular assembly of suitable chalcogenide anion species in the presence of metals. Recently, the synthesis of mesostructured solids based on the adamantane  $[\text{Ge}_4\text{Q}_{10}]^{4-}$  (where Q = S, Se) clusters and metal ions (such as  $\text{Zn}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ) has been described. Depending on preparation conditions, these materials possess either a disordered three-dimensional framework structure or a regular hexagonal symmetry<sup>11–14</sup>. The energy bandgaps of these solids however are still too high for electronic applications.

We describe the design and synthesis of mesostructured materials based on the elementary tetrahedral Zintl anion  $[\text{SnSe}_4]^{4-}$ . This anion is a chemical and structural analogue of the  $\text{SiO}_4$  unit and we expect it to yield topologically similar structures. It is also smaller and heavier than the  $[\text{Ge}_4\text{S}_{10}]$  cluster and more likely to yield narrow-gap semiconductors. Spatially controlled assembly of  $[\text{SnSe}_4]^{4-}$  anions<sup>15</sup> in formamide with various divalent metals in the presence of cetylpyridinium molecules as the surfactant template resulted in the formation of new mesophases. These materials, indicated as  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$  (where  $1.0 < x < 1.3$ ; M =  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Cd}^{2+}$  and  $\text{Hg}^{2+}$ ), show uniform pore size and large diversity in pore organization, depending on the metal, and ranging from wormhole to hexagonal to cubic. The elemental composition  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$  was determined by energy-dispersive microprobe analysis (EDS), elemental C, H, N and thermogravimetric analysis (TGA). In all samples the ratio of Sn:Se was very close to 1:4, in agreement with the expected ratio for the tetrahedral  $[\text{SnSe}_4]^{4-}$  anions. Infrared spectroscopy confirmed the presence of cetylpyridinium ions in the mesostructured  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$  solids. A summary of the analytical and other data is shown in Table 1.

X-ray diffraction (XRD) patterns of the products show a strong, sharp peak at low scattering angles corresponding to (100) reflections (Fig. 1). The  $d_{100}$  values, which represent the pore–pore separation, vary with the transition metal used, ranging between 35 Å and 40 Å (see Table 2). The XRD patterns also show a broad higher-order peak ( $2\theta \approx 4\text{--}5$  degrees), corresponding to overlapping higher-order reflections (110) and (200) which can be indexed to a hexagonal lattice as observed by transmission electron

**Table 1** Elemental analysis, colours and bandgaps for mesostructured metal tin selenides

| Sample     | Sn:Se <sup>*</sup> | M:Se <sup>*</sup> | Percentage of C, H, N | M:Se <sup>†</sup> | Colour        | Bandgap (eV) |
|------------|--------------------|-------------------|-----------------------|-------------------|---------------|--------------|
| Mn         | 1.01               | 0.90              | 44.71, 7.10, 2.25     | 1.03              | Orange        | 2.0          |
| Fe         | 0.91               | 1.22              | 38.40, 6.23, 2.05     | 1.28              | Dark brown    | 1.4          |
| Co         | 0.98               | 0.94              | 48.35, 6.64, 2.36     | 1.19              | Dark brown    |              |
| Zn (cubic) | 0.99               | 0.86              | 40.28, 6.17, 2.18     | 1.20              | Yellow-orange | 2.5          |
| Zn (hex)   | 0.96               | 0.85              | 41.58, 6.44, 2.29     | 1.15              | Yellow        | 2.5          |
| Cd         | 0.92               | 1.23              | 36.90, 6.01, 2.28     | 1.23              | Yellow        | 2.4          |
| Hg         | 1.03               | 1.09              | 34.65, 5.52, 1.92     | 1.20              | Dark orange   | 2.2          |

\* Based on energy-dispersive spectroscopy (EDS) results. Semiquantitative microprobe analyses (EDS) were performed on a JEOL JSM-6400 scanning electron microscope (SEM) equipped with a Noran EDS system. Data acquisition was performed several times in different areas of the samples using an accelerating voltage of 25 kV and 60-s accumulation time. Quoted values were obtained from an average of four measurements.

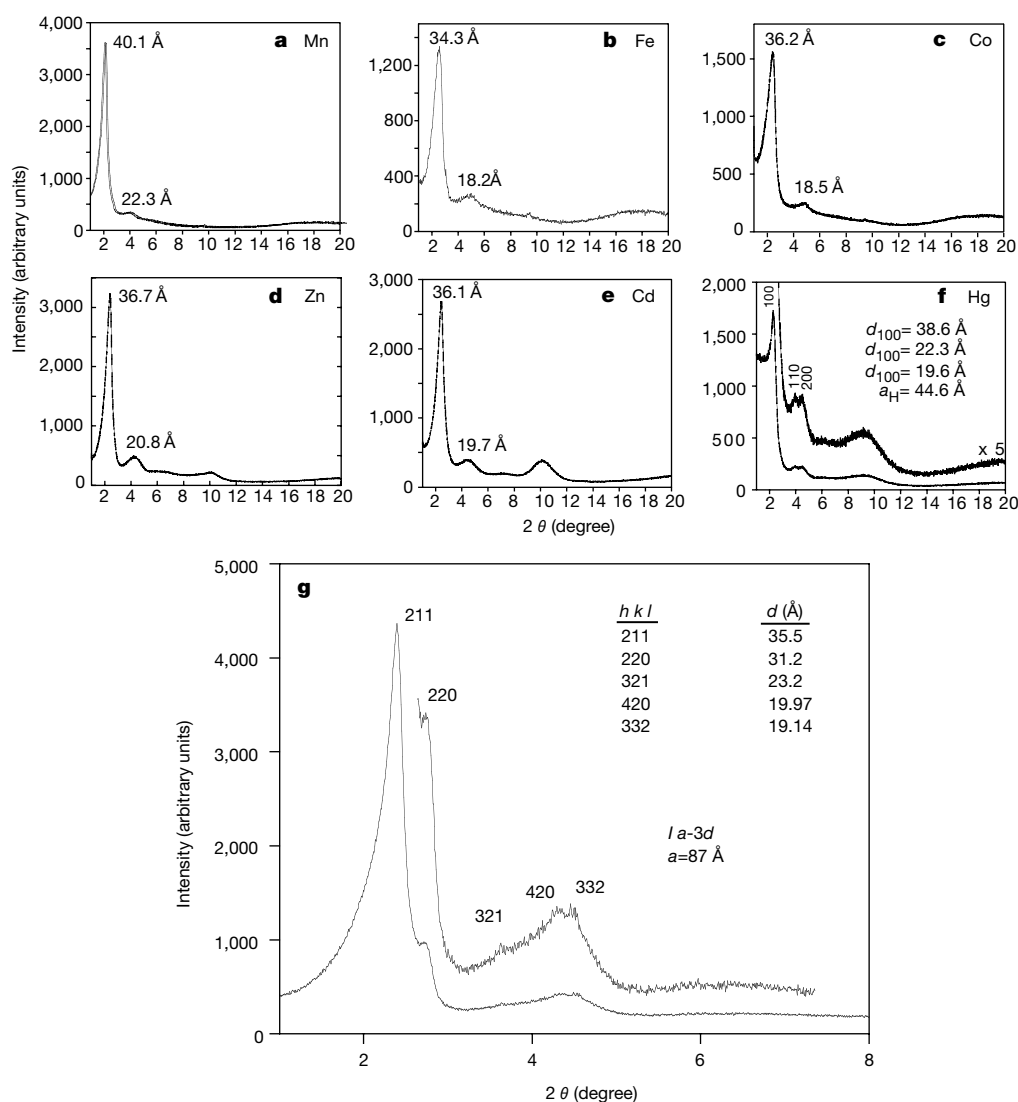
† Calculated values according to the formula  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$ , based on C, H, N results. C, H, N analyses were performed with Perkin Elmer Series II CHNS/O Analyzer 2400.

microscopy (TEM) (see below). The (110) and (200) reflections are well resolved in  $(\text{CP})_{4-2x}\text{Hg}_x\text{SnSe}_4$ , owing to the higher degree of order and longer coherence lengths, as observed in typical XRD patterns of the hexagonal mesostructured silica MCM-41<sup>3,4</sup>.

It is remarkable that under the same experimental conditions the structure of the Zn analogue,  $c\text{-(CP)}_{4-2x}\text{Zn}_x\text{SnSe}_4$ , is in fact cubic, as judged from its diffraction pattern in which up to five Bragg peaks can be resolved and indexed to a  $1a\text{-}3d$  symmetry (Fig. 1g). Because of the presence of some disorder, evident in the TEM images shown below, the higher-order Bragg peaks observed for  $c\text{-(CP)}_{4-2x}\text{Zn}_x\text{SnSe}_4$  are weak. To the best of our knowledge this is the first report of a cubic non-oxide mesostructured material. In the  $\text{CP}/\text{Zn}/\text{SnSe}_4$  system the surfactant concentration plays a crucial role in directing the mesostructure. For example, when the concentration is reduced by 50%, the product,  $h\text{-(CP)}_{4-2x}\text{Zn}_x\text{SnSe}_4$ , has a hexagonal pore arrangement with a corresponding XRD pattern (Fig. 1d). We note here that even among the silicas the cubic version MCM-48 is considered to be more interesting and more important for practical applications than the hexagonal versions MCM-41. Cubic MCM-48 is more difficult to make.

The local mesopore organization in  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$  is readily

observed by TEM. It is clear from all images that the pore sizes are in the range of 22–26 Å and are uniform throughout the particles. Figure 2a and b shows characteristic images of  $h\text{-CP-Zn}/\text{SnSe}_4$  and  $\text{CP-Hg}/\text{SnSe}_4$  parallel to the pore channel axis where the hexagonal mesostructure is clearly visible. Although the local hexagonal order of the pores is distinctly observed in these samples, the TEM micrographs reveal that the degree of organized domains varies over wide regions of the particles. This is similar to the mesoporous aluminophosphates UHM-X<sup>16</sup>, where the particles show local hexagonal pore packing alternating with wormhole domains. It is also consistent with the observed XRD patterns as discussed above. Figure 2c and d shows a view of  $\text{CP-Hg}/\text{SnSe}_4$  and  $\text{CP-Mn}/\text{SnSe}_4$  perpendicular to the pore axis (in the [110] direction). Long, straight parallel tunnels are apparent in these images and the observed inter pore distances are in good agreement with those obtained from the XRD patterns, Fig. 1. In fact, the agreement between XRD and TEM for all materials is excellent. Figure 2e and f shows representative TEM images of particles of  $c\text{-CP-Zn}/\text{SnSe}_4$  and  $\text{CP-Cd}/\text{SnSe}_4$  in which the pore organization is different from those shown in Fig. 2a and b. These images reveal cubic-like pore packing along more disordered



**Figure 1** X-ray scattering of mesostructured chalcogenides. **a**,  $\text{CP-Mn}/\text{SnSe}_4$ . The second peak at 22.3 Å suggests that the pores are organized in finite but ordered domains. **b**,  $\text{CP-Fe}/\text{SnSe}_4$ ; **c**,  $\text{CP-Co}/\text{SnSe}_4$ ; **d**,  $h\text{-CP-Zn}/\text{SnSe}_4$ ; **e**,  $\text{CP-Cd}/\text{SnSe}_4$ . **f**,  $\text{CP-Hg}/\text{SnSe}_4$ . The pore organization in this material into hexagonal domains is more

extensive. **g**, X-ray diffraction pattern of cubic  $c\text{-CP-Zn}/\text{SnSe}_4$ , with indexing. This pattern is similar to those observed in cubic silica MCM-48 systems. Powder X-ray data were collected with a Rigaku Rotaflex powder X-ray diffractometer equipped with Ni-filtered Cu K $\alpha$  radiation and operating at 45 kV and 100 mA.



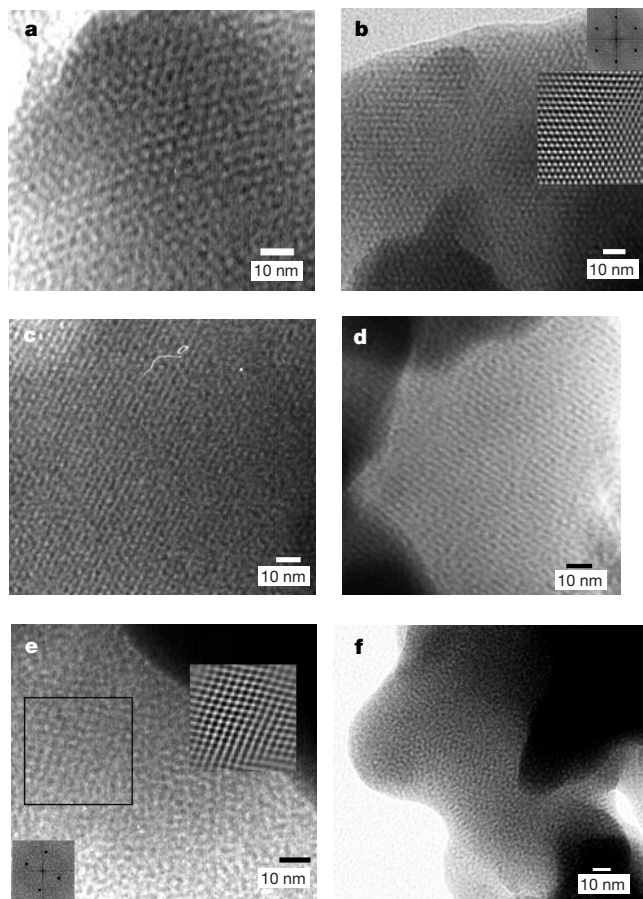
regions for CP-Zn/SnSe<sub>4</sub> and nearly wormhole-like arrangement for CP-Cd/SnSe<sub>4</sub>.

The presence of the tetrahedral [SnSe<sub>4</sub>]<sup>4-</sup> anion in these materials is supported by <sup>119</sup>Sn Mössbauer spectroscopy, a powerful experimental technique for the identification and characterization of the Sn local environment and formal oxidation state. Figure 3 shows Mössbauer spectra for K<sub>4</sub>SnSe<sub>4</sub>, CP-Mn/SnSe<sub>4</sub>, CP-Fe/SnSe<sub>4</sub>, CP-Zn/SnSe<sub>4</sub>, CP-Cd/SnSe<sub>4</sub> and CP-Hg/SnSe<sub>4</sub>. The corresponding isomer shift,  $\delta$ , and quadrupole splitting,  $\Delta E_q$ , parameters are very similar in all compounds and characteristic of tetrahedral Sn<sup>4+</sup> centres<sup>17</sup> (see Fig. 3 legend). The appearance of quadrupole splitting in the spectra of the mesophases suggests that the local lattice symmetry of the Sn<sup>4+</sup> centre is lower compared to that in K<sub>4</sub>SnSe<sub>4</sub>. This is attributed to binding of the metal centres which lowers the tetrahedral symmetry. The spectra also reveal two slightly different Sn environments, suggesting at least two different binding modes for SnSe<sub>4</sub>.

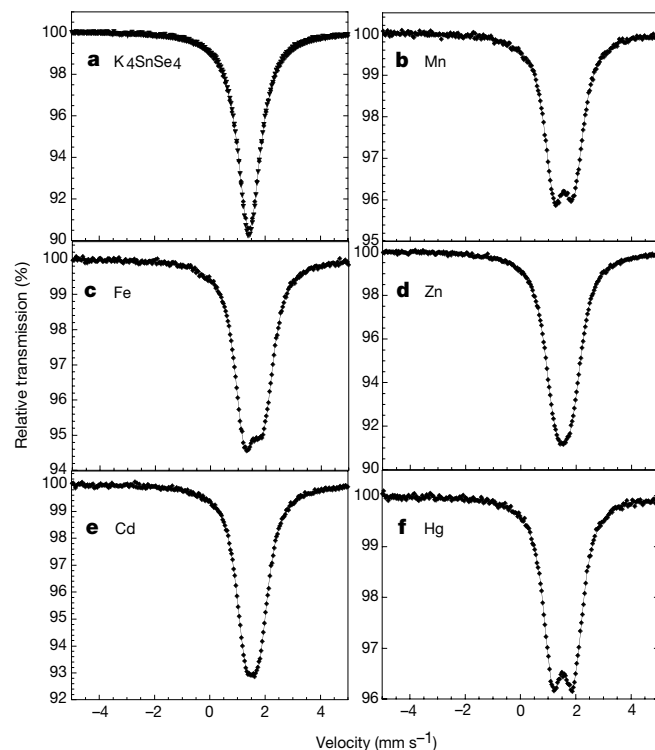
The thermal stability of (CP)<sub>4-2x</sub>M<sub>x</sub>SnSe<sub>4</sub> was investigated by TGA and pyrolysis mass spectrometry (MS). (Pyrolysis MS analyses were carried out on a Hewlett Packard Series II 5890 coupled with a

mass spectrometer. Samples were heated at 20 °C min<sup>-1</sup> and the volatile products were ionized by electron ionization.) See Fig. 4a for characteristic TGA curves of CP-Co/SnSe<sub>4</sub>, CP-Zn/SnSe<sub>4</sub>, CP-Cd/SnSe<sub>4</sub> and CP-Hg/SnSe<sub>4</sub>. The compounds show no appreciable weight loss up to 150 °C. Between 150 and 400 °C weight loss occurs as a result of surfactant decomposition. The XRD powder patterns of the residue correspond to a mixture of crystalline SnSe and MSe. The samples CP-Cd/SnSe<sub>4</sub> and CP-Hg/SnSe<sub>4</sub> clearly show a two-step decomposition process. The first occurs between 150 and 290 °C and the second between 290 and 400 °C, leading to a mixture of MSe and SnSe. The weight loss observed in the first step corresponds to quantitative release (~11%) of the pyridyl head group from the surfactant and this was verified with pyrolysis MS analysis. We note that the volatile product released between 300–400 °C was identified to be CH<sub>3</sub>(CH<sub>2</sub>)<sub>13</sub>CH = CHSeH, showing that the metal chalcogenide framework is attacked during this process. Unfortunately, this form of ‘calcination’ is not the correct one, as the complete removal of the surfactant species results in framework collapse.

The optical absorption properties of the solids were investigated with diffuse-reflectance solid-state ultraviolet–visible/near-infrared spectroscopy. In general, chalcogenide materials have considerably narrower energy bandgaps than oxides (for example, silicates) and this property makes them attractive for optoelectronic applications. The (CP)<sub>4-2x</sub>M<sub>x</sub>SnSe<sub>4</sub> compounds possess well defined, sharp optical absorptions associated with bandgap transitions in the energy range 1.4–2.5 eV (see Table 1 and Fig. 4b). The very sharp electronic transitions in the Zn, Cd and Hg analogues suggest that their bandgaps may possibly be direct. In addition, the bandgap narrows in going from the lighter to the heavier element in the isoelectronic



**Figure 2** Transmission electron microscope images of mesostructured chalcogenides. **a**, *h*-CP-Zn/SnSe<sub>4</sub> looking down the pore channel axis; **b**, CP-Hg/SnSe<sub>4</sub> looking down the pore channel axis. Inset in upper right, optical Fourier transform of image and corresponding filtered image. The straight parallel nature of the pore tunnels extends over >1,500 Å. **c**, CP-Hg/SnSe<sub>4</sub> perpendicular to the channel axis; **d**, CP-Mn/SnSe<sub>4</sub> perpendicular to the channel axis. **e**, *c*-CP-Zn/SnSe<sub>4</sub>. The cubic arrangement is shown in the boxed area. Insets in lower left and upper right, optical Fourier transform of boxed area, and corresponding filtered image. **f**, CP-Cd/SnSe<sub>4</sub>. Disordered regions with wormhole arrangement. TEM samples were prepared by suspending the precipitate in ether, then casting on holey carbon grid. High-resolution TEMs were acquired with a JEOL 120CX instrument equipped with a CeB<sub>6</sub> filament and operating at 120 kV. The TEM images are typical and representative of the samples under observation.



**Figure 3** <sup>119</sup>Sn Mössbauer spectra (85 K) with derived  $\delta/\Delta E_q$  parameters. **a**, K<sub>4</sub>SnSe<sub>4</sub> (standard) 1.42/0 mm s<sup>-1</sup>. **b**, CP-Mn/SnSe<sub>4</sub>, 1.54/0.86 (52%) mm s<sup>-1</sup> and 1.54/0.52 (48%) mm s<sup>-1</sup>. **c**, CP-Fe/SnSe<sub>4</sub>, 1.55/0.94 (50%) mm s<sup>-1</sup> and 1.55/0.42 (46%) mm s<sup>-1</sup>. **d**, CP-Zn/SnSe<sub>4</sub>, 1.52/0.97 (44%) mm s<sup>-1</sup> and 1.52/0.39 (56%) mm s<sup>-1</sup>. **e**, CP-Cd/SnSe<sub>4</sub>, 1.53/0.71 (50%) mm s<sup>-1</sup> and 1.53/0.33 (50%) mm s<sup>-1</sup>. **f**, CP-Hg/SnSe<sub>4</sub>, 1.54/1.13 (20%) and 1.53/0.62 (80%) mm s<sup>-1</sup>. The spectra clearly show that the Sn atoms are in the +4 oxidation state. The isomer shifts are referenced to BaSnO<sub>3</sub> at room temperature.

series. The Mn and Fe analogues possess the lowest energy gaps at  $\sim 1.4$  and  $2.0$  eV, values similar to those of CdTe and GaP semiconductors<sup>18</sup>.

The surfactant-templated supramolecular assembly of the simple tetrahedral Zintl  $[\text{SnSe}_4]^{4-}$  anion with transition metals leads to the formation of mesostructured semiconducting selenide materials of the general formulae  $(\text{CP})_{4-2x}\text{M}_x\text{SnSe}_4$  (where  $\text{M} = \text{Mn}, \text{Fe}, \text{Co}, \text{Zn}, \text{Cd}, \text{Hg}$ ). The pores in these systems are uniformly sized and show local hexagonal order as observed by TEM.

The  $c\text{-(CP)}_{4-2x}\text{M}_x\text{SnSe}_4$  represents the first example of cubic,

non-oxide mesophase. Ten years after the discovery of mesostructured cubic silica (MCM-48), the discovery of a semiconducting chalcogenide analogue is an important milestone. The reasons for this and the advantages over previous hexagonal systems are as follows. Cubic symmetry in a porous nanostructured semiconductor provides the possibility of creating the negative image of a quantum dot crystal, that is, an array of ordered 'quantum dot' nanocrystals, also of cubic symmetry. Such a material, instead of presenting ordered semiconductor nanocrystals separated by continuous dielectric space, presents continuous semiconductor space separated by dielectric voids. It can most easily be thought of as the geometric complement or 'inside-out version' of an array of ordered 'quantum dot' nanocrystals. These types of materials have been visualized in the nanoengineering physics literature<sup>1</sup> and have been termed 'exosemiconductors'<sup>2</sup> or 'quantum antidots'. The relationship between these two types of nanostructures has yet to be demonstrated. This is partly because currently there are no 'quantum antidot'-type materials.  $c\text{-(CP)}_{4-2x}\text{M}_x\text{SnSe}_4$  is a substantial step towards creating them. Cubic symmetry in this case allows for experimentation aimed at answering fundamental questions, such as: can exosemiconductors be quantized?

The present experiments establish that it is possible to construct organically templated structures with heavier  $\text{SiO}_4$ -analogues to produce semiconducting solids with well defined mesopores. Therefore they open new prospects for the design and construction of electronically active, porous chalcogenide solids and other non-oxide systems in general. These materials should help us to understand how ordered mesostructured materials form and may also enable unusual shape-selective electronically or optically driven processes. In particular, cubic nanostructured semiconductors provide an isotropic bi-continuous medium in which the pore system of the semiconductor framework can potentially be accessed by guest molecules from all directions. Three-dimensional guest diffusion in and out of the material can give rise to faster processes associated with whatever material function is being used. This, for example, could be molecule-discriminating semiconductor photocatalysis occurring in the interior of the material, rather than what is now available on the surface of the semiconductor. A hexagonal system with its unidirectional pore organization does not permit pore-pore communication and could frustrate guest diffusion. We anticipate cubic nanostructured semiconductors may be tailored in several ways, through the composition of the framework, arrangement of pores, even through the electronic properties of the organic amphiphiles, for optoelectronic, photosynthetic and photocatalytic applications and membrane probes. Further work involving other chalcogenide building blocks, such as  $[\text{GeSe}_4]^{4-}$ ,  $[\text{GeTe}_4]^{4-}$ ,  $[\text{SnTe}_4]^{4-}$ ,  $[\text{SbSe}_3]^{3-}$  and  $[\text{SbTe}_3]^{3-}$  is in progress.  $\square$

## Methods

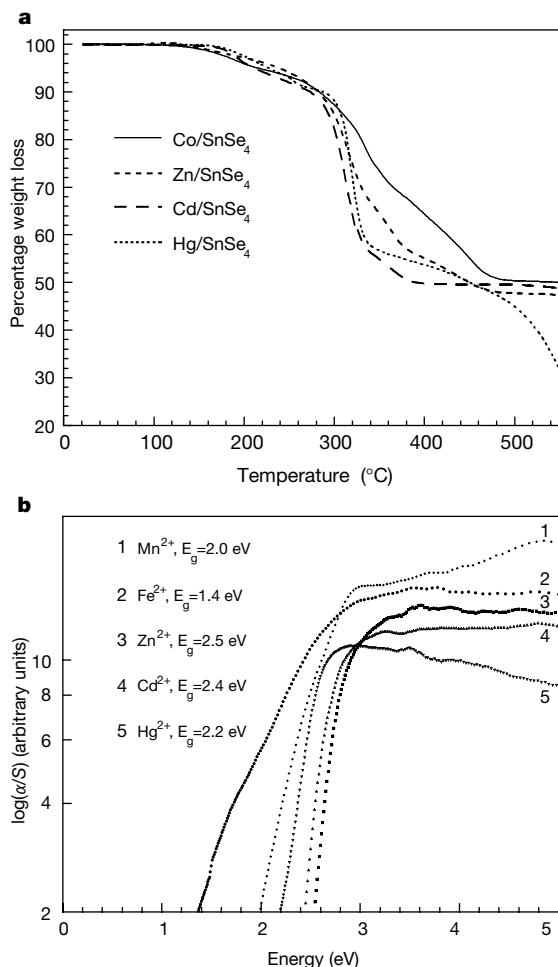
The starting materials were  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ ,  $\text{FeSO}_4 \cdot 6\text{H}_2\text{O}$ ,  $\text{Co}(\text{Ac})_2 \cdot 4\text{H}_2\text{O}$ ,  $\text{Zn}(\text{Ac})_2 \cdot 2\text{H}_2\text{O}$ ,  $\text{Cd}(\text{NO}_3)_2$  and  $\text{Hg}(\text{Ac})_2$  (where  $\text{Ac} = \text{CH}_3\text{COO}$ ). All reactions were carried out under  $\text{N}_2$  atmosphere inside a glove box.

## General procedure

In a typical preparation, a solution of  $4.00$  g ( $9.94$  mmol) of surfactant ( $\text{CPBr} \cdot \text{H}_2\text{O}$ ) in  $20$  ml of formamide was heated at  $75^\circ\text{C}$  for a few minutes forming a clear solution. To this solution,  $0.59$  g ( $1.00$  mmol) of  $\text{K}_4\text{SnSe}_4$  was added and the mixture was stirred for  $30$  min, forming a clear deep red solution. To this a solution of  $1.00$  mmol of the metal salt in  $10$  ml of formamide was added slowly, using a pipette. A precipitate formed immediately and the mixture was stirred while warm for  $20$  h. The products were isolated by suction filtration and washed with large amounts of warm formamide and water. The solids were dried under vacuum overnight. The yields of the orange (Mn), dark brown (Fe, Co), yellow (Zn, Cd) and dark orange (Hg) products were  $\sim 80\%$  in all cases, based on  $\text{K}_4\text{SnSe}_4$  as the limiting reagent. For  $(\text{CP})_{4-2x}\text{Zn}_x\text{SnSe}_4$  the above preparation gives a cubic phase. The hexagonal phase of this system requires  $2.00$  g ( $4.97$  mmol) of surfactant.

Received 12 September 2000; accepted 7 February 2001.

1. Weiss, D. *et al.* Quantized periodic-orbits in large antidot arrays. *Phys. Rev. Lett.* **70**, 4118–4121 (1993).



**Figure 4** Thermogravimetric data and optical absorption spectra. **a**, TGA curves of CP-Co/SnSe<sub>4</sub>, CP-Zn/SnSe<sub>4</sub>, CP-Cd/SnSe<sub>4</sub> and CP-Hg/SnSe<sub>4</sub>, under nitrogen flow. Heating rate was  $10^\circ\text{C min}^{-1}$ . Data were obtained with a Shimadzu TGA-50 thermal analyser. **b**, Optical absorption spectra of selected mesostructured chalcogenides. (Shimadzu UV-3101PC double-beam, double-monochromator spectrophotometer in the wavelength range  $200\text{--}2,500$  nm). Absorption  $(\alpha/S)$  data were calculated from the reflectance data using the Kubelka–Munk function:  $\alpha/S = (1 - R)^2/2R$ , where  $R$  is the reflectance at a given wavenumber,  $\alpha$  is the absorption coefficient, and  $S$  is the scattering coefficient.

**Table 2**  $d$ -values for mesostructured metal tin selenides

| Sample     | $d_{100}$ values (Å) | Second-order scattering $d$ values (Å) |
|------------|----------------------|----------------------------------------|
| Mn         | 40.1                 | 22.3                                   |
| Fe         | 34.3                 | 18.2                                   |
| Co         | 36.2                 | 18.5                                   |
| Zn (cubic) | $d_{211} = 35.5$     | 31.2 (see Fig. 1g)                     |
| Zn (hex)   | 36.7                 | 20.8                                   |
| Cd         | 36.1                 | 19.7                                   |
| Hg         | 38.6                 | $d_{110} = 22.3, d_{200} = 19.6$       |

2. Ozin, G. A. Nanomaterials—endosemiconductors and exosemiconductors. *Adv. Chem. Ser.* **245**, 335–371 (1995).
3. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
4. Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid-crystal templates. *J. Am. Chem. Soc.* **114**, 10834–10843 (1992).
5. Ying, Y. Y., Mehnert, C. P. & Wong, M. S. Synthesis and applications of supramolecular-templated mesoporous materials. *Angew. Chem. Int. Edn Engl.* **38**, 56–77 (1999).
6. Ciesla, U., Schacht, S., Stucky, G. D., Unger, K. K. & Schuth, F. Formation of a porous zirconium oxo phosphate with a high surface area by a surfactant-assisted synthesis. *Angew. Chem. Int. Edn Engl.* **35**, 541–543 (1996).
7. Attard, G. S., Goltner, C. G., Corker, J. M., Henke, S. & Templar, R. H. Liquid-crystal templates for nanostructured metals. *Angew. Chem. Int. Edn Engl.* **36**, 1315–1317 (1997).
8. Braun, P. V., Osenar, P. & Stupp, S. I. Semiconducting superlattices templated by molecular assemblies. *Nature* **380**, 325–328 (1996).
9. Li, H. L. *et al.* Supertetrahedral sulfide crystals with giant cavities and channels. *Science* **283**, 1145–1147 (1999).
10. Bowes, C. L. & Ozin, G. A. Self-assembling frameworks: Beyond microporous oxides. *Adv. Mater.* **8**, 13–18 (1996).
11. MacLachlan, M. J., Coombs, N. & Ozin, G. A. Non-aqueous supramolecular assembly of mesostructured metal germanium sulphides from  $(\text{Ge}_4\text{S}_{10})^{4-}$  clusters. *Nature* **397**, 681–684 (1999).
12. Wachhold, M. *et al.* Mesostructured non-oxidic solids with adjustable worm-hole shaped pores: M-Ge-Q (Q = S, Se) frameworks based on tetrahedral  $[\text{Ge}_4\text{Q}_{10}]^{4-}$  clusters. *Adv. Mater.* **12**, 85–91 (2000).
13. Rangan, K. K., Billinge, S. J. L., Petkov, V., Heising, J. & Kanatzidis, M. G. Aqueous mediated synthesis of mesostructured manganese germanium sulfide with hexagonal order. *Chem. Mater.* **11**, 2629–2632 (1999).
14. Wachhold, M. *et al.* Mesostructured metal germanium sulfide and selenide materials based on the tetrahedral  $[\text{Ge}_4\text{S}_{10}]^{4-}$  and  $[\text{Ge}_4\text{Q}_{10}]^{4-}$  units: Surfactant templated three-dimensional disordered frameworks perforated with worm holes. *J. Solid State Chem.* **152**, 21–36 (2000).
15. Klepp, K. O.  $\text{Na}_4\text{SnSe}_4$  and  $\text{K}_4\text{SnSe}_4$ , new selenostannates with discrete anions. *Z. Naturforsch. B* **47**, 411–417 (1992).
16. Zhao, D. Y., Luan, Z. H. & Kevan, L. Synthesis of thermally stable mesoporous hexagonal aluminophosphate molecular sieves. *Chem. Commun.* 1009–1010 (1997).
17. Lippens, P. E. Interpretation of  $^{119}\text{Sn}$  Mössbauer isomer shifts in complex tin chalcogenides. *Phys. Rev. B* **60**, 4576–4586 (1999).
18. Pankove, J. I. *Optical Processes in Semiconductors* (Dover, New York, 1971).

#### Acknowledgements

This work made use of the SEM and TEM facilities of the Center for Advanced Microscopy at Michigan State University. Financial support from the National Science Foundation is gratefully acknowledged.

Correspondence and requests for materials should be addressed to M.G.K. (e-mail: kanatzid@cem.msu.edu).

## An alternative approach to establishing trade-offs among greenhouse gases

Alan S. Manne\* & Richard G. Richels†

\*Stanford University, Stanford, California 94305, USA

†Electric Power Research Institute (EPRI), PO Box 10412, Palo Alto, California 94303, USA

The Kyoto Protocol permits countries to meet part of their emission reduction obligations by cutting back on gases other than  $\text{CO}_2$  (ref. 1). This approach requires a definition of trade-offs among the radiatively active gases. The Intergovernmental Panel on Climate Change has suggested global warming potentials for this purpose<sup>2</sup>, which use the accumulated radiative forcing of each gas by a set time horizon to establish emission equivalence. But it has been suggested that this approach has serious shortcomings: damages or abatement costs are not considered<sup>3–10</sup> and the choice of time horizon for calculating cumulative radiative force is critical, but arbitrary<sup>5</sup>. Here we describe an alternative framework for determining emission equivalence between radiatively active

gases that addresses these weaknesses. We focus on limiting temperature change and rate of temperature change, but our framework is also applicable to other objectives. For a proposed ceiling, we calculate how much one should be willing to pay for emitting an additional unit of each gas. The relative prices then determine the trade-off between gases at each point in time, taking into account economical as well as physical considerations. Our analysis shows that the relative prices are sensitive to the lifetime of the gases, the choice of target and the proximity of the target, making short-lived gases more expensive to emit as we approach the prescribed ceiling.

Although the Kyoto Protocol encompasses a number of radiatively active gases, the assessment of compliance costs has focused almost exclusively on the costs of reducing carbon dioxide ( $\text{CO}_2$ ) emissions. This is because  $\text{CO}_2$  is, by far, the most important man-made gas<sup>2</sup>. Also, until recently, few economic models have been able to conduct comprehensive multi-gas analyses<sup>10</sup>. Moreover, the quality of data pertaining to other greenhouse gases is poor (both spatially and intertemporally). Nevertheless, focusing exclusively on  $\text{CO}_2$  will bias mitigation cost estimates and lead to policies that are unnecessarily costly.

A number of gases have now been identified as having a positive effect on radiative forcing. We consider the three that are thought to be the most important:  $\text{CO}_2$ , methane ( $\text{CH}_4$ ) and nitrous oxide ( $\text{N}_2\text{O}$ ) (ref. 2). We also consider the cooling effect of sulphate aerosols, but exclude the 'second basket' of greenhouse gases included in the Kyoto Protocol. These are the hydrofluorocarbons (HFCs), the perfluorocarbons (PFCs) and sulphur hexafluoride ( $\text{SF}_6$ ). We believe that these omissions do not alter the key insights of our analysis.

Table 1 shows alternative global warming potentials (GWPs) for the three gases of interest. The index is defined as the cumulative radiative forcing between the present and some time in the future caused by a unit mass of gas emitted now, expressed relative to that of  $\text{CO}_2$  (ref. 2). Clearly, the choice of time horizon is critical. Unfortunately, as pointed out in ref. 5, there is no justification for choosing one time horizon over another. GWPs are a purely physical measure. They depend neither on damages nor mitigation costs. For a discussion of the uncertainties involved in calculating GWPs, see ref. 11.

What might then constitute a more defensible index for establishing trade-offs among gases? Ideally, the index would be the outcome of an analysis that minimizes the discounted present value of damages and mitigation costs. Unfortunately, we lack at present the necessary knowledge to specify the shapes of the damage functions and to assign values to many categories of impacts. We have therefore used an approach in which the ceiling is intended to reflect a political judgement as to what constitutes a prudent limit. These limits are based upon expectations regarding the damages associated with particular ceilings. The appropriate trade-offs among gases are then determined through a cost-effectiveness analysis.

Unlike the calculation of GWPs, the proposed approach incorporates the marginal cost of abating each greenhouse gas. This can have important implications for the trade-offs among gases. The more expensive it is to abate a particular gas, the smaller the role of that gas in a multi-gas reduction portfolio. Relying completely on

**Table 1 GWPs as a function of alternative time horizons**

| Species        | Chemical formula     | Lifetime (years) | Global warming potential (years) |     |     |
|----------------|----------------------|------------------|----------------------------------|-----|-----|
|                |                      |                  | 20                               | 100 | 500 |
| Carbon dioxide | $\text{CO}_2$        | 50–200           | 1                                | 1   | 1   |
| Methane        | $\text{CH}_4$        | 12               | 56                               | 21  | 6.5 |
| Nitrous oxide  | $\text{N}_2\text{O}$ | 120              | 280                              | 310 | 170 |

Data from ref. 2.



purely physical measures to determine the trade-offs among gases ignores this potentially important consideration.

The proposed approach is consistent with the UN Framework Convention on Climate Change<sup>12</sup>, which has as its ultimate goal the “stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system”. As there are many ways of achieving this goal, the Framework Convention goes on to state that “policies and measures to deal with climate change should be cost-effective so as to ensure global benefits at the lowest possible cost” (ref. 12).

The analysis is based on a computable general equilibrium (CGE) model called MERGE. (For a description of the model and its key inputs, see <http://www.stanford.edu/group/MERGE/>). The version used in the present analysis integrates submodels that provide a reduced-form description of the energy sector, the economy, emissions, concentrations and temperature change. MERGE is disaggregated across both space and time. The globe is divided into nine geopolitical regions. The choice of time horizon is contingent upon the issue under study. For the current application, it is necessary that the proposed ceiling be binding within the time horizon of the model. Here, 200 years are sufficient.

It is particularly important for the present analysis to be able to calculate the prices of the various greenhouse gases. That is, for a given ceiling, these prices express how much one should be willing to pay to emit an additional ton of each gas. The trade-offs among gases are then based on the relative prices of each gas as determined by its contribution to the ceiling.

Here we focus on temperature change (both absolute and decadal). However, we stress that the approach is applicable to any ceiling to which the greenhouse gases contribute. In addition to identifying an economically efficient temperature profile for satisfying a particular temperature goal, MERGE also identifies the corresponding concentration profiles and emissions trajectories.

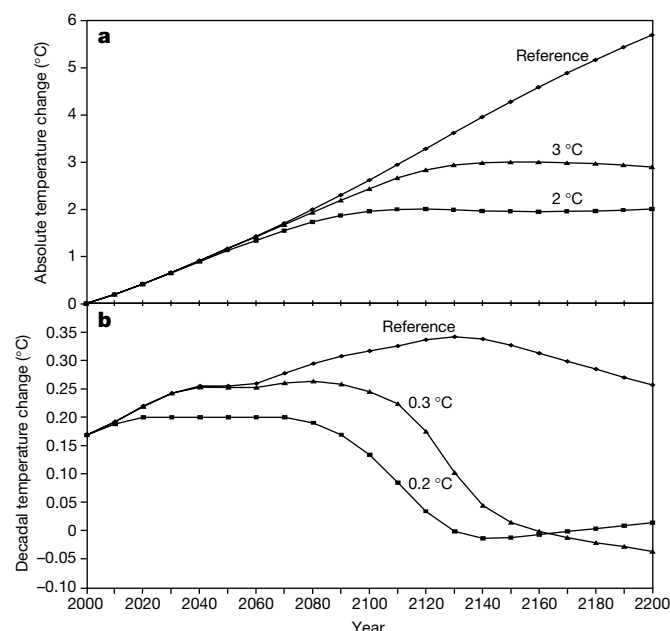
We begin by assuming that the goal of climate policy is to limit the

future increase in mean global temperature. Using MERGE, we identify an economically efficient strategy for staying within the prescribed ceiling. Figure 1a shows the temperature trajectory for a reference case and for ceilings of 2°C and 3°C. These trajectories incorporate the cooling effects of sulphate aerosols. Such effects are assumed to depend largely on local and regional air quality considerations and to be independent of global climate policy.

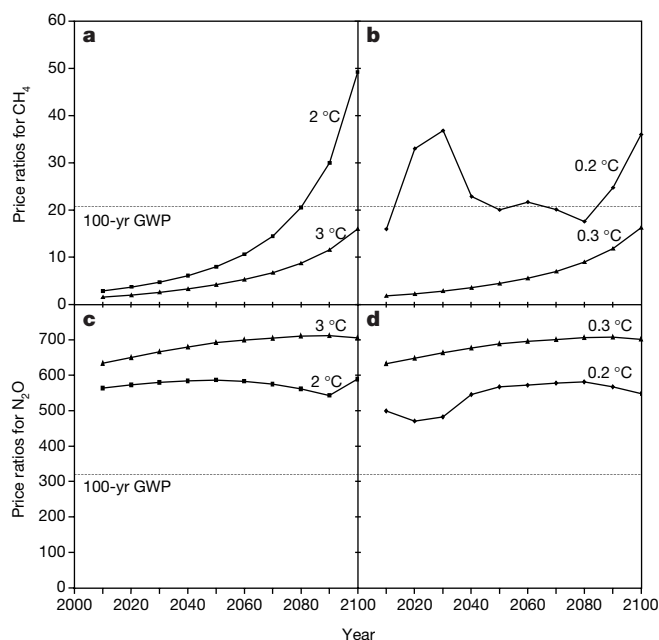
Some have suggested that we should be concerned with both absolute temperature change and the rate of temperature change<sup>13</sup>. To explore this possibility, we also examine situations in which an additional constraint is imposed on the two temperature scenarios. We limit the allowable increase during a single decade to 10% of the total allowable increase. That is, decadal temperature change is limited to 0.2°C and 0.3°C, respectively. Figure 1b shows that with a ceiling of 2°C on absolute temperature change, there are decades during the twenty-first century where the limit on the rate of temperature change would be binding.

Figure 2a and c shows the price of the other two gases relative to that of carbon dioxide for temperature ceilings of 2°C or 3°C. It also shows the 100-year GWPs for each gas. We note that the relative prices vary over time. This is particularly so for CH<sub>4</sub>. With its relatively short lifetime, a ton emitted in the early decades of the twenty-first century will have only a small effect on temperature in the late twenty-first century. As we approach the temperature ceiling, it becomes more worthwhile to reduce the emissions of a short-lived gas, relative to that of a long-lived gas. With a lifetime more nearly commensurate with that of CO<sub>2</sub>, the price ratios for N<sub>2</sub>O are less volatile. However, like CH<sub>4</sub>, they are sensitive to the choice of ceiling. Limiting the temperature increase to 2 rather than 3°C produces a different set of relative prices.

Figure 2b and d shows the implications of combining the temperature ceiling with a rate of change constraint. We note that the price ratios for CH<sub>4</sub> reflect what we observed earlier: the closer we are to the temperature constraint, the more valuable CH<sub>4</sub> becomes. This appears to be true whether the constraint is only



**Figure 1** Absolute temperature change and decadal temperature change under alternative scenarios. For details regarding the composition of the reference case, see our website <http://www.stanford.edu/group/MERGE/>. **a**, Economically efficient temperature profiles for limiting absolute temperature change to 2 and 3°C. **b**, Decadal temperature change when a rate of change constraint is also imposed. Here, we limit the allowable increase for a single decade to 10% of the total allowable increase.



**Figure 2** The prices of CH<sub>4</sub> and N<sub>2</sub>O relative to that of CO<sub>2</sub> under alternative constraints on absolute and decadal temperature change. **a**, **c**, Prices of CH<sub>4</sub> and N<sub>2</sub>O relative to that of CO<sub>2</sub> when the ceiling is on absolute temperature change. **b**, **d**, The corresponding results when a rate of change constraint is added. GWP, global warming potential.

on absolute temperature, or on absolute and the decadal rate of temperature change.

Finally, to confirm the importance of the relatively short lifetime of CH<sub>4</sub> in determining its role in the portfolio of abatement measures, we conducted a hypothetical experiment in which the lifetime of CH<sub>4</sub> was doubled. As one would expect, we found that CH<sub>4</sub> prices rise more slowly as we approach a ceiling. This is because it takes more time for the impact of a reduction to be realized.

Earlier, we noted two problems with GWPs: the failure to incorporate damages and abatement costs, and the arbitrary choice of time horizon for calculating cumulative radiative forcing. Here we highlight two additional problems. GWPs assume that the trade-off ratios remain constant over time and are independent of the ultimate goal. Clearly, neither of these assumptions makes economic sense. The relative prices are a function of both the target and the proximity to the target. It is illogical to suppose that this is a case of "one size fits all"; yet this is precisely what is suggested by the IPCC in recommending the use of 100-year GWPs.

Unlike GWPs, the alternative we propose extends beyond purely physical considerations in calculating trade-offs among gases. Expectations about damages influence the choice of ceiling. Abatement costs influence the relative roles of the various gases in the portfolio of abatement actions. For example, we found that the higher the costs of abating CH<sub>4</sub>, the larger the role of the other gases in a multi-gas reduction portfolio.

The approach also provides a clear definition of what constitutes an appropriate time horizon for the analysis. And it permits the necessary flexibility for the price ratios to vary both with the choice of and proximity to the target.

We believe that the proposed approach represents a more defensible way for making trade-offs among gases. Clearly, large uncertainties remain both with regard to damages (and hence, in what constitutes an appropriate ceiling) and abatement costs. Extensive sensitivity analysis is needed to determine which factors are most influential in determining the price ratios and thus where reducing uncertainty will have the highest pay-off. Nevertheless, uncertainty need not lead to paralysis. The issue is how to make the best near-term choices in the face of the many long-term uncertainties. As with other aspects of the climate debate, this requires a willingness to learn and make mid-course corrections. Price ratios can be designed on the basis of the best available information about both the physical system and the energy-economic system. As better information emerges, the price ratios can be modified accordingly. □

Received 7 June 2000; accepted 14 February 2001.

1. UNFCCC *Kyoto Protocol to the United Nations Framework Convention on Climate Change* (UNEP/INC/98/2, Information Unit for Conventions, UNEP, Geneva, 1998); (<http://www.unfccc.int/resource/docs/convkp/kpeng.pdf>).
2. Schimel, D. *et al.* in *Climate Change 1995: The Science of Climate Change* 65–131 (eds Houghton, J. T. *et al.*) (Cambridge Univ. Press, Cambridge, 1996).
3. Eckaus, R. Comparing the effects of greenhouse gas emissions on global warming. *Energy J.* **13**, 25–34 (1992).
4. Reilly, J. M. & Richards, K. R. Climate change damage and trace gas index issue. *Environ. Res. Econ.* **3**, 41–61 (1993).
5. Schmalensee, R. Comparing greenhouse gases for policy purposes. *Energy J.* **14**, 245–255 (1993).
6. Fankhauser, S. *Valuing Climate Change – the Economics of the Greenhouse* (EarthScan, London, 1995).
7. Kandlikar, M. The relative role of trace gas emissions in greenhouse abatement. *Energy Policy* **23**, 879–883 (1995).
8. Hammitt, J. *et al.* A welfare based index for assessing environmental effects of greenhouse gas emissions. *Nature* **381**, 301–303 (1996).
9. Tol, R. *et al.* *Methane Emission Reduction: An Application of Funds 1–44* (Institute for Environmental Studies, Vrije Universiteit, The Netherlands, 1999).
10. Reilly, J. *et al.* Multi-gas assessment of the Kyoto Protocol. *Nature* **401**, 549–555 (1999).
11. Wuebbles, D. J., Jain, A. K., Patten, K. O. & Grant, K. E. Sensitivity of direct global warming potentials to uncertainties. *Clim. Change* **29**, 265–297 (1995).
12. Climate Change Secretariat. *United Nations Framework Convention on Climate Change 6* (UNEP/IUC, Geneva, 1992).

13. Peck, S. C. & Teisberg, T. J. Optimal carbon emission trajectories when damages depend on the rate or level of global warming. *Clim. Change* **28**, 289–315 (1994).

## Acknowledgements

We thank C. Gerlach and R. Parkin for research assistance. We have benefited from discussions with J. Aldy, J. Edmonds, L. Goulder, J. Reilly, R. Tol, J. Weyant, T. Wigley, L. Williams and T. Wilson. Funding was provided by EPRI.

Correspondence and requests for materials should be addressed to R. R. (e-mail: rrichels@epri.com).

## Coupled major and trace elements as indicators of the extent of melting in mid-ocean-ridge peridotites

Eric Hellebrand\*, Jonathan E. Snow\*, Henry J. B. Dick† & Albrecht W. Hofmann\*

\*Max-Planck-Institut für Chemie, Postfach 3060, D-55020 Mainz, Germany

†Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Rocks in the Earth's uppermost sub-oceanic mantle, known as abyssal peridotites, have lost variable but generally large amounts of basaltic melt, which subsequently forms the oceanic crust<sup>1,2</sup>. This process preferentially removes from the peridotite some major constituents such as aluminium, as well as trace elements that are incompatible in mantle minerals (that is, prefer to enter the basaltic melt), such as the rare-earth elements<sup>3,4</sup>. A quantitative understanding of this important differentiation process has been hampered by the lack of correlation generally observed between major- and trace-element depletions in such peridotites. Here we show that the heavy rare-earth elements in abyssal clinopyroxenes that are moderately incompatible are highly correlated with the Cr/(Cr + Al) ratios of coexisting spinels. This correlation deteriorates only for the most highly incompatible elements—probably owing to late metasomatic processes. Using electron- and ion-microprobe data from residual abyssal peridotites collected on the central Indian ridge, along with previously published data, we develop a quantitative melting indicator for mantle residues. This procedure should prove useful for relating partial melting in peridotites to geodynamic variables such as spreading rate and mantle temperature.

We studied minerals in 22 spinel peridotite samples, dredged from seven locations along the central Indian ridge (CIR) and its fracture zones. These samples are petrographically residual, lacking plagioclase and crosscutting magmatic veins. All samples are harzburgites except for one orthopyroxene-bearing dunite and two lherzolites. Clinopyroxenes (cpxs) have rare-earth-element patterns that are depleted in light rare-earth elements (LREEs), and all cover the most depleted range of the global abyssal peridotite spectrum<sup>3,4</sup>. Whereas the heavy rare-earth elements (HREEs) have a limited range in concentration (~2–10 times chondritic), the LREE abundances vary by more than an order of magnitude (Table 1).

To evaluate the relationships between major and trace elements, we selected all published global abyssal peridotites for which both mineral major-element and cpx trace-element data are available. This includes samples from three locations on the southwest Indian ridge (SWIR) and two on the American–Antarctic ridge (AAR)<sup>3,4</sup>. In addition, we used data from drill cores at Hess deep (East Pacific Rise)<sup>5</sup> and the MARK area (Mid-Atlantic Ridge near Kane fracture zone)<sup>6</sup>, giving a reasonable sampling of global abyssal peridotite occurrences.

First we tested which trace elements correlate with major-element indicators of partial melting. The most common of these is the chromium number (Cr#) in spinel<sup>7,8</sup> (Fig. 1). (Cr# is the molar Cr/(Cr + Al) ratio.) In the CIR peridotites spinel Cr# values range from 0.17 to 0.57, which is 80% of the global abyssal peridotite spectrum. There is a well defined correlation between moderately incompatible elements, such as HREEs dysprosium, erbium and ytterbium, with spinel Cr# (Fig. 1). All elements were fitted with an exponential line, as the fractional melting equation is an exponential function. Fits were poor (see  $r^2$  values in Fig. 1) for highly incompatible elements (for example, Ce, Sm, Ce/Yb) and good for less incompatible elements (for example, Dy, Yb). Only highly incompatible elements, it seems, are highly decoupled from major-element mineral chemistry. Cerium, the most incompatible element in this selection, shows the largest variation at a given Cr#. Both strong Ce depletion at low Cr# and large variations towards higher concentrations at higher Cr# suggest that analytical uncertainty at low concentrations is not the only responsible factor. A variety of processes can explain LREE enrichments in mantle rocks<sup>9–13</sup>.

Correlation coefficients calculated between trace elements and three major-element melting indicators (Cr# in spinel,  $Al_2O_3$  and Mg number (Mg#) in cpx) confirm the visual impression of Fig. 1. (Mg# is the molar [Mg/(Mg + Fe)] ratio.) This is summarized in Fig. 2, where the experimentally determined cpx/melt distribution coefficient ( $K_d^{cpx/l}$ ) of each measured trace element<sup>14</sup> is plotted versus  $r^2$  for the correlation with spinel Cr#. The HREEs (Dy, Er, Yb) in residual cpx are well correlated with spinel Cr# ( $r^2 \approx 0.9$ ) and the aluminium content of cpx ( $r^2 \approx 0.75$ , for  $n = 57$ ). Correlations with middle rare-earth elements (MREEs) (Sm, Eu) and Ti are markedly lower but still significant. LREEs (La, Ce and Nd), Sr and Zr are only weakly or not at all correlated. The Mg# of residual silicate phases is often regarded as a powerful monitor of melting, which increases with increasing degree of melting. However, Fig. 2 shows that Mg# in cpx does not correlate significantly with other melting indicators. This is probably due to its small absolute range in residual oceanic mantle silicates (between 0.90 and 0.93 for cpx),

and is thus largely analytical.

The good correlation between the major elements and the moderately incompatible trace elements (MREE and HREE) indicates that both are highly useful tracers of partial melting in the mantle. Furthermore, the presence and absence of correlations provide us with a guide on how to interpret abyssal peridotite mineral data. Considering that REE abundances in cpx can be used for quantitative melting modelling<sup>3</sup>, the good correlation of spinel Cr# with HREE concentration in cpx enables us to develop an empirical equation that describes the extent of melting ( $F$ ) as a function of spinel Cr#. As a limiting case, we use pure fractional melting in the spinel stability field to calculate the degree of melting after starting conditions given in ref. 3, except for a lower initial cpx mode of 0.14. Mineral/melt partition coefficients were taken from ref. 14. The measured concentrations of Dy, Er and Yb each yield a degree of melting  $F$ , which is plotted versus the Cr# of the associated spinel (Fig. 3). A logarithmic fit to all these calculated points yields an equation for the degree of melting (in per cent) as a function of the spinel Cr#:  $F = 10 \ln(Cr\#) + 24$ , calibrated for spinel Cr# values between 0.10 and 0.60 (thick line in Fig. 3). A similar equation has been derived for mantle rocks in Bay of Islands ophiolite<sup>15</sup>.

In order to assess how robust this equation is, we applied different melting models using otherwise identical starting parameters. The average model melting curve assuming batch melting (dashed line) and critical melting with 1% residual melt porosity (thin line) are also shown in Fig. 3 (the model data points themselves are not shown). The small difference between pure fractional and critical melting with a relatively high melt porosity confirms that the moderately incompatible HREEs are not sensitive to changes in melt porosity within the melting column, as long as cpx remains in the residue. We note that Fig. 3 by itself does not indicate whether melt extraction is by fractional or batch mechanisms.

As an independent test, spinel Cr#–liquid fraction pairs from 10-kbar batch melting experiments<sup>16</sup> are plotted as filled symbols in Fig. 3. These are consistent with our calculated melting curve as long as cpx is present in the residue. By treating major elements

**Table 1 REE concentrations in cpx and spinel Cr numbers of CIR peridotites**

| Sample              | Location     | <i>n</i> | Lab. | La     | Ce     | Nd    | Sm    | Eu    | Dy    | Er    | Yb    | Spinel Cr# |
|---------------------|--------------|----------|------|--------|--------|-------|-------|-------|-------|-------|-------|------------|
| CIRCE 93-2          | CIR 12° S    | 2        | MPI  | 0.014  | 0.028  | 0.06  | 0.12  | 0.06  | 0.75  | 0.64  | 0.72  | 0.32       |
|                     |              | 2        | WHOI |        |        | 0.12  | 0.14  | 0.07  | 0.64  | 0.51  | 0.49  |            |
| CIRCE 93-4          | CIR 12° S    | 2        | MPI  | 0.003  | 0.013  | 0.11  | 0.11  | 0.05  | 0.89  | 0.63  | 0.74  | 0.31       |
|                     |              | 2        | WHOI |        | 0.016  | 0.13  | 0.14  | 0.06  | 0.85  | 0.60  | 0.58  |            |
| CIRCE 93-7          | CIR 12° S    | 2        | MPI  | 0.015  | 0.095  | 0.50  | 0.46  | 0.21  | 0.81  | 0.44  | 0.53  | 0.39       |
| ANTP 126-HD2        | AFZ (13° S)  | 2        | MPI  | 0.009  | 0.049  | 0.04  | 0.07  | 0.04  | 0.73  | 0.61  | 0.66  | 0.22       |
| ANTP 126-HD5        | AFZ (13° S)  | 2        | MPI  | 0.010  | 0.045  | 0.06  | 0.07  | 0.04  | 0.73  | 0.61  | 0.66  | 0.22       |
|                     |              | 2        | WHOI |        | 0.041  | 0.09  | 0.09  | 0.05  | 0.61  | 0.47  | 0.57  |            |
| ANTP 126-D5         | AFZ (13° S)  | 2        | MPI  | 0.005  | 0.020  | 0.05  | 0.13  | 0.06  | 0.94  | 0.71  | 0.84  | 0.22       |
|                     |              | 2        | WHOI |        | 0.020  | 0.08  | 0.12  | 0.08  | 0.76  | 0.54  | 0.62  |            |
| ANTP 134-HD3        | VFZ (8° S)   | 2        | MPI  | 0.001  | 0.014  | 0.24  | 0.34  | 0.14  | 1.22  | 0.84  | 0.86  | 0.27       |
|                     |              | 2        | WHOI |        | 0.016  | 0.25  | 0.25  | 0.12  | 0.86  | 0.62  | 0.56  |            |
| ANTP 134-HD4        | VFZ (8° S)   | 3        | MPI  | 0.002  | 0.036  | 0.26  | 0.27  | 0.13  | 1.09  | 0.73  | 0.83  | 0.31       |
|                     |              | 3        | WHOI |        | 0.035  | 0.31  | 0.30  | 0.12  | 0.78  | 0.55  | 0.57  |            |
| ANTP 134-HD7        | VFZ (8° S)   | 4        | WHOI |        | 0.025  | 0.28  | 0.25  | 0.12  | 0.85  | 0.59  | 0.56  | 0.32       |
| ANTP 134-HD8        | VFZ (8° S)   | 1        | MPI  | 0.002  | 0.024  | 0.19  | 0.19  | 0.08  | 0.93  | 0.54  | 0.58  | 0.30       |
|                     |              | 3        | WHOI |        | 0.033  | 0.29  | 0.26  | 0.14  | 0.87  | 0.56  | 0.62  |            |
| ME33/2: 31GTV       | GRH (25° S)  | 2        | MPI  | 0.002  | 0.004  | 0.04  | 0.06  | 0.03  | 0.61  | 0.43  | 0.58  | 0.38       |
| SO92: 60GTV         | GRH (25° S)  | 1        | MPI  | 0.005  | 0.017  | 0.03  | 0.09  | 0.04  | 0.60  | 0.43  | 0.58  | 0.38       |
| SO92: 74GTV         | GRH (25° S)  | 3        | MPI  | 0.002  | 0.003  | 0.015 | 0.05  | 0.03  | 0.47  | 0.39  | 0.49  | 0.38       |
| ANTP 84-HD11        | MCFZ (18° S) | 2        | MPI  | b.d.   | 0.002  | 0.005 | 0.009 | 0.003 | 0.29  | 0.29  | 0.38  | 0.33       |
| ANTP 87-HD5         | MCFZ (18° S) | 2        | MPI  | 0.002  | 0.002  | b.d.  | 0.03  | 0.02  | 0.41  | 0.46  | 0.45  | 0.32       |
| ANTP 87-HD9         | MCFZ (18° S) | 2        | MPI  | 0.003  | 0.006  | 0.015 | 0.03  | 0.03  | 0.38  | 0.34  | 0.45  | 0.38       |
| ANTP 89-HD1         | MCFZ (18° S) | 2        | MPI  | 0.002  | 0.013  | 0.21  | 0.30  | 0.13  | 1.32  | 0.83  | 0.74  | 0.29       |
| ANTP 89-HD2         | MCFZ (18° S) | 2        | MPI  | 0.002  | 0.006  | 0.06  | 0.16  | 0.10  | 1.21  | 0.84  | 0.75  | 0.26       |
| ANTP 89-HD5         | MCFZ (18° S) | 3        | MPI  | 0.001  | 0.005  | 0.13  | 0.23  | 0.13  | 1.36  | 0.94  | 0.87  | 0.17       |
| ANTP 89-HD8         | MCFZ (18° S) | 2        | MPI  | 0.003  | 0.006  | 0.12  | 0.21  | 0.13  | 1.32  | 0.97  | 0.78  | 0.20       |
| ANTP 89-HD15        | MCFZ (18° S) | 4        | MPI  | 0.001  | 0.008  | 0.17  | 0.21  | 0.13  | 1.24  | 0.92  | 0.72  | 0.29       |
| ANTP 89-HD17        | MCFZ (18° S) | 4        | MPI  | 0.014  | 0.013  | 0.012 | 0.06  | 0.04  | 0.64  | 0.57  | 0.65  | 0.28       |
| REE detection limit |              |          | MPI  | 0.0006 | 0.0007 | 0.003 | 0.004 | 0.001 | 0.002 | 0.003 | 0.001 |            |

REE concentrations are given in  $\mu\text{g g}^{-1}$ . Trace elements were analysed in Mainz (MPI) by SIMS (recently upgraded Cameca IMS-3f), applying the energy filtering technique of ref. 27. Operating conditions of the Mainz ion microprobe were 12.5 kV accelerating potential and 20 nA primary beam current. Standard glass GOR132 was used as an external standard<sup>28</sup>. Average internal precision for the data is better than 10% for HREE, and better than 15% for Eu. The internal precision for LREE and MREE is concentration-dependent, and the error increases with decreasing counting rate. For La and Ce the internal precision varies around ~20–30% for concentrations of  $10 \text{ ng g}^{-1}$  and ~50% for concentrations of  $2 \text{ ng g}^{-1}$ . For Nd and Sm the internal precision is better than 20% at concentrations higher than  $100 \text{ ng g}^{-1}$  and ~40% for concentrations around  $20 \text{ ng g}^{-1}$ . Some samples were also measured at the Cameca IMS-3f ion microprobe at Woods Hole Oceanographic Institution (WHOI) as noted. Complete petrography and mineral chemistry is given in ref. 29. GRH, Green Rock hill; MCFZ, Marie Celeste fracture zone; *n*, number of analyses; b.d., below detection limit. REE detection limit is  $6 \times$  background.



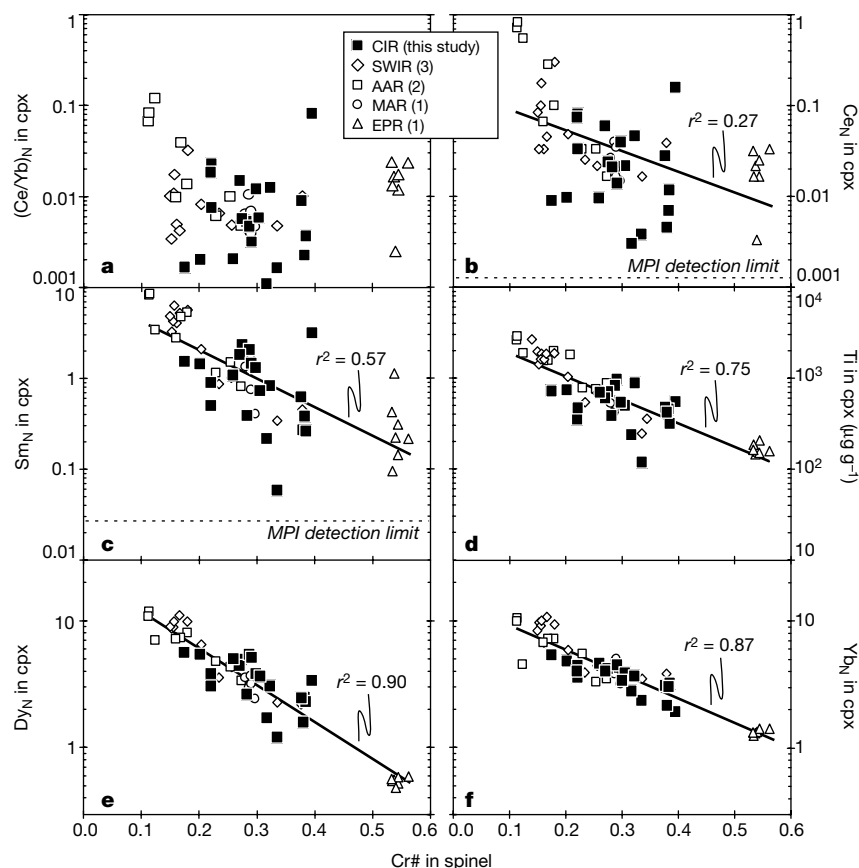
chromium and aluminium as trace elements, spinel compositions can be modelled directly to a first approximation (partition coefficients are calculated from the melting experiments of ref. 16). Batch and fractional melting models yield similar spinel compositions, both of which agree with our results. By contrast, in our models  $F$  is calculated from incompatible trace elements Dy, Er and Yb in cpx, which show considerable differences between fractional and batch mechanisms (Fig. 3). The experimental data plot above our fractional melting curve at low  $F$ . This is because the starting composition chosen for this particular experiment is relatively rich in chromium, resulting in a significantly higher initial spinel Cr# than observed in the most fertile abyssal peridotites, and in a melting trend that is shifted towards higher Cr# values at a given melt fraction.

Using both major- and trace-element melting indicators, some inferences can begin to be drawn regarding the geodynamic controls on melting globally. First, if only non-hotspot peridotites are considered, the degree of melting appears to correlate with spreading rate<sup>17</sup>. Although moderately incompatible trace elements confirm this inference, this relationship must be treated with great caution. Hess deep and the MARK area are the only locations on their respective ridges yet investigated for both major and trace elements, and the comparatively well investigated SWIR, AAR and CIR show large compositional variation. Therefore, regional-scale 'forcing functions' deserve more attention.

A geodynamic control on the degree of melting, one which is superimposed on the effect of the spreading rate, is the vicinity to a

transform offset. Most abyssal peridotites are collected at fracture zones. Mantle rocks from Hess deep, on the other hand, were generated in the centre of a spreading segment far from a transform fault<sup>5</sup>, which may explain their extreme depletion. In our database, only three other locations are not situated directly at a major fracture zone (MARK area, Green Rock hill on CIR 25°S, and CIRCE 93 samples on CIR axis 12°S). The last two are among the most depleted samples from the CIR, indicating that all other CIR fracture zone samples may have been affected by a transform fault effect<sup>18,19</sup>. Our quantitative tool of Fig. 3 can now be used to estimate the degree of melting for oceanic mantle rocks where often only spinel data are available. The MARK area is of particular importance, because a large set of mantle melting residues contains evidence for this transform fault effect<sup>20</sup>. Peridotites directly at Kane fracture zone have low Cr# values of around 0.17, a value that systematically increases to 0.41 in highly depleted residues near the ridge centre (Fig. 3). Applying our melting function under fractional melting conditions, we estimate the minimal magnitude of the transform fault effect. The most depleted peridotite corresponds to 15% melting and the most fertile sample to 5%. Thus, the minimal difference in the degree of melting, which may be caused by deepening of the final depth of melting towards the transform fault, is 10% (that is, a factor of three).

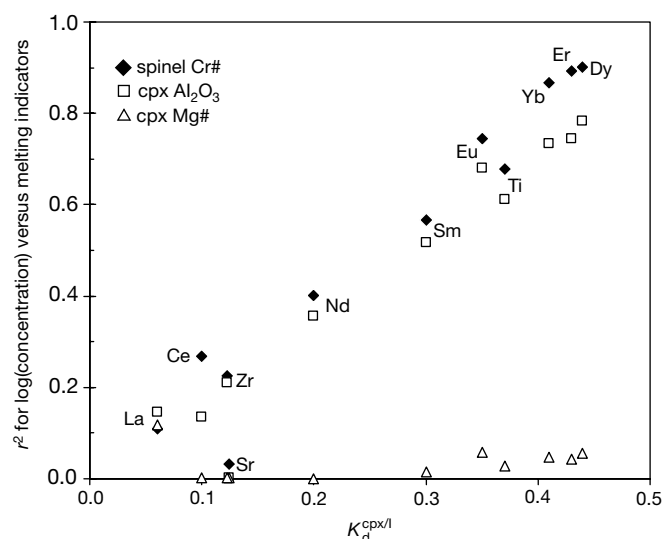
Further, the vicinity of a hotspot is believed to increase the degree of melting. The strong trace-element depletion in the Bouvet fracture zone peridotites has been attributed<sup>3</sup> to this phenomenon. Comparing these depleted Bouvet fracture zone samples to the



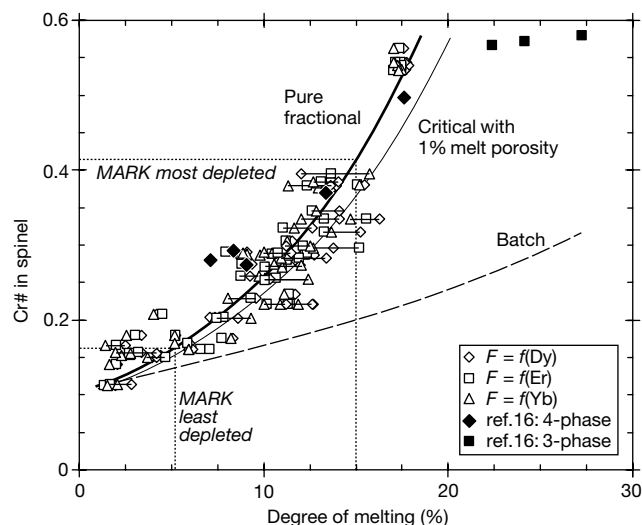
**Figure 1** Mineral major- and trace-element systematics in residual abyssal peridotites. **a**, Cr number (Cr#) in spinel versus  $(\text{Ce/Yb})_N$  in cpx. (Subscript N indicates normalization to chondritic values.) Theoretical melting models predict that these parameters are negatively correlated, but they are in fact highly decoupled. **b-f**, Plots of Cr# in spinel versus incompatible element concentration in coexisting cpx  $\text{Ce}_N$  (**b**),  $\text{Sm}_N$  (**c**), Ti ( $\mu\text{g g}^{-1}$ )

(**d**),  $\text{Dy}_N$  (**e**) and  $\text{Yb}_N$  (**f**) show progressively tighter correlations expressed by the correlation coefficient ( $r^2$ ). Numbers in parentheses are number of locations for that portion of mid-ocean ridge. All points are sample averages. Data from refs 3–6, 23 and Table 1; chondrite normalizing values from ref. 24.

fertile Islas Orcadas fracture zone peridotites away from the Bouvet hotspot provides an estimate of the additional hotspot-imposed melting, because of the identical spreading rate at the SWIR and the location at a fracture zone. For the SWIR, the additional degree of melting related to the vicinity of the Bouvet hotspot is 8%. However, the Bouvet fracture zone mantle rocks are not uniquely depleted.



**Figure 2** Plot of  $r^2$  versus the cpx/liquid distribution coefficient. Values for  $r^2$  were obtained by plotting melting indicators (spinel Cr#, cpx  $\text{Al}_2\text{O}_3$  content, and cpx Mg number, Mg#) versus the logarithmic concentration of each trace element. All values of  $K_d^{\text{cpx/l}}$  are from ref. 14, except Sr (ref. 25). HREEs are well correlated with spinel Cr#, and trace elements become progressively decoupled from spinel Cr# with increasing incompatibility. There is no significant correlation between cpx Mg# and any of the trace-element concentrations.



**Figure 3** Calculated degree of melting versus measured Cr# of spinel. Degrees of melting are calculated from concentrations of Dy, Er and Yb in cpx, yielding three independent degrees of melting for each spinel Cr#. All individual model points are only shown for pure fractional melting (open symbols). The exponential fit (thick curve) is based on average degree of melting for each sample. The same procedure yields melting trajectories for batch melting (dashed curve) and critical melting with 1% residual melt porosity (thin curve) using equations after ref. 26. The degree of fractional melting  $F$  (in per cent) as a function of Cr# can be expressed as  $F = 10 \ln(\text{Cr\#}) + 24$  for Cr# values between 0.1 and 0.6. Most and least depleted Cr# compositions from the MARK area are shown. Experimental data shown agree with the melting trend for cpx-bearing residues shown as filled diamonds<sup>16</sup>.

Harzburgites from Green Rock hill have nearly identical major- and trace-element compositions to the Bouvet samples. Yet, no nearby hotspot has been detected in the area of Rodrigues triple junction. It therefore seems that the geodynamic combination of near-hotspot, transform fault and slow-spreading ridge can produce similar residual mantle rocks as a hotspot-free, non-transform, intermediate-spreading setting. However, if the global abyssal peridotite data set is filtered for samples from hotspots and transform faults, peridotite compositions from the few locations left do show a correlation with spreading rate<sup>21</sup>. Mantle temperature variations should in principle also produce strong effects on the degree of melting in the peridotite residues<sup>22</sup>.

Another question is whether the observed correlations shown in Fig. 1 may also be inherited from the mantle source before melting. The calculations that are based on these observations assume that the pre-melting Cr# values and HREE concentrations are uniform. It is also possible that the two parameters were already similarly correlated before the most recent (mid-ocean-ridge basalt) melting event. If this were true, we would expect that the calculated degree of melting should not be correlated with tectonic parameters such as spreading rate. Further, this would imply that the systematic depletion trend of the MARK area peridotites<sup>20</sup> is fortuitous and not directly related to a transform fault effect. Such inherited depletion might be correlated with abundances of radiogenic isotopes. Future studies that focus on the relationship between melting and regional tectonics will help resolve this issue. □

Received 31 August 2000; accepted 31 January 2001.

- Hess, H. H. In *Petrologic Studies* (eds Engel, A. E. J., James, H. L. & Leonard, B. F.) 599–620 (Burlington Volume, Geological Society of America, Boulder, 1962).
- Bonatti, E. Ultramafic rocks from the Mid-Atlantic ridge. *Nature* **219**, 363–364 (1968).
- Johnson, K. T. M., Dick, H. J. B. & Shimizu, N. Melting in the oceanic upper mantle; an ion microprobe study of diopsides in abyssal peridotites. *J. Geophys. Res.* **95**, 2661–2678 (1990).
- Johnson, K. T. M. & Dick, H. J. B. Open system melting and temporal and spatial variation of peridotite and basalt at the Atlantis II fracture zone. *J. Geophys. Res.* **97**, 9219–9241 (1992).
- Dick, H. J. B. & Natland, J. H. Late-stage melt evolution and transport in the shallow mantle beneath the East Pacific Rise. *Proc. ODP Sci. Res.* **147**, 103–134 (1996).
- Ross, K. & Elthon, D. Extreme incompatible trace-element depletion of diopside in residual mantle from south of the Kane F.Z. *Proc. ODP Sci. Res.* **153**, 277–284 (1997).
- Dick, H. J. B., Fisher, R. L. & Bryan, W. B. Mineralogical variability of the uppermost mantle along mid-ocean ridges. *Earth Planet. Sci. Lett.* **69**, 88–106 (1984).
- Michael, P. J. & Bonatti, E. Peridotite composition from the North Atlantic; regional and tectonic variations and implications for partial melting. *Earth Planet. Sci. Lett.* **73**, 91–104 (1985).
- Bodinier, J. L., Vasseur, G., Vernières, J., Dupuy, C. & Fabries, J. Mechanisms of mantle metasomatism: geochemical evidence from the Lherz orogenic peridotite. *J. Petrol.* **31**, 597–628 (1990).
- Kelmen, P. B., Shimizu, N. & Salters, V. J. M. Extraction of mid-ocean-ridge basalt from the upwelling mantle by focused flow of melt in dunite channels. *Nature* **375**, 747–753 (1995).
- Seyler, M. & Bonatti, E. Regional-scale melt-rock interaction in Iherzolitic mantle in the Romanche fracture zone (Atlantic Ocean). *Earth Planet. Sci. Lett.* **146**, 273–287 (1997).
- Suhr, G. Melt migration under oceanic ridges: inferences from reactive transport modelling of upper mantle hosted dunites. *J. Petrol.* **40**, 575–599 (1999).
- Pearce, J. A., Barker, P. F., Edwards, S. J., Parkinson, I. J. & Leat, P. T. Geochemistry and tectonic significance of peridotites from the South Sandwich arc-basin system, South Atlantic. *Contrib. Mineral. Petrol.* **139**, 36–53 (2000).
- Suhr, G., Seck, H. A., Shimizu, N., Günther, D. & Jenner, G. Infiltration of refractory melts into the lowermost oceanic crust: evidence from dunite- and gabbro-hosted clinopyroxenes in the Bay of Islands Ophiolite. *Contrib. Mineral. Petrol.* **131**, 136–154 (1998).
- Batanova, V. G., Suhr, G. & Sobolev, A. V. Origin of geochemical heterogeneity in the mantle peridotites from the Bay of Islands Ophiolite, Newfoundland, Canada; ion probe study of clinopyroxenes. *Geochim. Cosmochim. Acta* **62**, 853–866 (1998).
- Baker, M. B. & Stolper, E. M. Determining the composition of high-pressure mantle melts using diamond aggregates. *Geochim. Cosmochim. Acta* **58**, 2811–2827 (1994).
- Niu, Y. & Hekinian, R. Spreading-rate dependence of the extent of mantle melting beneath ocean ridges. *Nature* **385**, 326–329 (1997).
- Bender, J. F., Langmuir, C. H. & Hanson, G. N. Petrogenesis of basalts from the Tamayo Region, East Pacific Rise. *J. Petrol.* **25**, 231–254 (1984).
- Natland, J. H. & Melson, W. G. Composition of basaltic glasses from the East Pacific Rise and Siqueros fracture zone, near 9 degrees N. *Init. Rep. DSDP* **54**, 705–724 (1980).
- Ghose, I., Cannat, M. & Seyler, M. Transform fault effect on mantle melting in the MARK area (Mid-Atlantic Ridge south of the Kane transform). *Geology* **24**, 1139–1142 (1996).
- Hellebrand, E., Snow, J. E. & Mühe, R. Mantle melting beneath the Gakkel Ridge (Arctic Ocean): abyssal peridotite spinel compositions. *Chem. Geol.* (in press).
- Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalts with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
- Dick, H. J. B. Abyssal peridotites, very slow spreading ridges and ocean ridge magmatism. *Geol. Soc. Spec. Publ.* **42**, 71–105 (1989).
- Anders, E. & Grevesse, N. Abundances of the elements: meteoric and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).

25. Hart, S. R. & Dunn, T. Experimental cpx/melt partitioning of 24 trace elements. *Contrib. Mineral. Petrol.* **113**, 1–8 (1993).
26. Sobolev, A. V. & Shimizu, N. Extremely depleted magmas and oceanic mantle permeability. *Dokl. Akad. Nauk* **326**, 354–360 (1992).
27. Shimizu, N., Semet, M. P. & Allegre, J. C. Geochemical applications of quantitative ion microprobe analysis. *Geochim. Cosmochim. Acta* **42**, 1321–1334 (1978).
28. Jochum, K. P. et al. The preparation and preliminary characterisation of eight geological MPI-DING standard reference glasses for in-situ microanalysis. *Geostandards Newsl.* **24**, 87–133 (2000).
29. Hellebrand, E., Snow, J. E. & Hofmann, A. W. Garnet-field melting and late-stage refertilization in 'residual' abyssal peridotites from the Central Indian Ridge. *J. Petrol.* (submitted).

#### Acknowledgements

We thank R. Fisher who originally collected these samples, P. Hoppe and A. Besmehn for their assistance on the Mainz ion probe, N. Shimizu and G. Layne on the WHOI ion probe, and G. Suhr for letting us use his critical melting algorithm. Comments by G. Suhr, A. Sobolev and J. Lassiter helped to improve the manuscript. We thank E. Klein and P. Asimov for formal reviews.

Correspondence and requests for materials should be addressed to E.H.  
(e-mail: ehelle@mpch-mainz.mpg.de).

## Complex causes of amphibian population declines

Joseph M. Kiesecker\*, Andrew R. Blaustein† & Lisa K. Belden†

\* Department of Biology, The Pennsylvania State University,  
208 Mueller Laboratory, University Park, Pennsylvania 16802-5301, USA  
† Department of Zoology, Oregon State University, Corvallis, Oregon 97331, USA

Amphibian populations have suffered widespread declines and extinctions in recent decades. Although climatic changes, increased exposure to ultraviolet-B (UV-B) radiation and increased prevalence of disease have all been implicated at particular localities<sup>1–6</sup>, the importance of global environmental change remains unclear. Here we report that pathogen outbreaks in amphibian populations in the western USA are linked to climate-induced changes in UV-B exposure. Using long-term observational data and a field experiment, we examine patterns among interannual variability in precipitation, UV-B exposure and infection by a pathogenic oomycete, *Saprolegnia ferax*. Our findings indicate that climate-induced reductions in water depth at oviposition sites have caused high mortality of embryos by increasing their exposure to UV-B radiation and, consequently, their vulnerability to infection<sup>1</sup>. Precipitation, and thus water depth/UV-B exposure, is strongly linked to El Niño/Southern Oscillation cycles, underscoring the role of large-scale climatic patterns involving the tropical Pacific<sup>7</sup>. Elevated sea-surface temperatures in this region since the mid-1970s, which have affected the climate over much of the world<sup>8</sup>, could be the precursor for pathogen-mediated amphibian declines in many regions<sup>1,3,4,9</sup>.

*Saprolegnia ferax* outbreaks due to increased UV-B exposure have been identified as a cause of high amphibian embryo mortality in the Pacific Northwest<sup>1</sup>. Attempts to link mortality patterns that are consistent with species declines have proposed that these patterns are connected to ozone depletion<sup>2</sup>. Although moderate decreases in total ozone over mid-latitudes have been reported<sup>10–13</sup>, changes in ambient UV-B radiation resulting from ozone depletion will probably occur against a background of natural and anthropogenic environmental changes that may act synergistically to alter exposure to UV-B radiation. In some cases, climate change can be more effective than stratospheric ozone depletion in increasing the exposure of aquatic organisms to biologically effective UV-B radiation, whereas in other cases climate change may act synergistically

with ozone depletion to increase UV-B exposure<sup>14,15</sup>. Amphibians, particularly those that breed in shallow montane lakes and ponds, may be quite susceptible to climate-induced changes in UV-B exposure. Amphibian embryos developing in montane lakes and ponds are often exposed to direct sunlight<sup>16,17</sup>; however, the overlying water column may attenuate UV-B radiation. Where precipitation is reduced, associated reduction in water depth at oviposition sites may enhance UV-B exposure. In the Pacific Northwest of the United States, where there are numerous reports of increased amphibian embryo mortality<sup>1,2,18–21</sup>, precipitation patterns are closely linked to El Niño/Southern Oscillation (ENSO) cycles<sup>22</sup>. Thus, the increase in frequency and magnitude of El Niño events after the step-like warming of the tropical Pacific<sup>7,8</sup> may have raised the incidence and severity of *S. ferax* outbreaks by increasing the extent to which embryos are exposed to sunlight in shallow water.

To examine the relationship between pathogen-mediated embryo mortality and climate, we quantified mortality in relation to water depth at natural oviposition sites in the context of interannual variation in precipitation and the Southern Oscillation Index (SOI). We used experimentation and observation to test three predictions regarding *S. ferax*-associated embryo mortality of western toads, *Bufo boreas*: (1) mortality associated with *S. ferax* infections at natural oviposition sites is related to the water depth in which embryos develop; (2) water depth at an oviposition site is a function of variability in precipitation associated with ENSO cycles; and (3) outbreaks of *S. ferax* infection observed in shallow water are mediated by exposure to UV-B radiation.

Breeding activity of *B. boreas* was monitored at several locations in the Oregon Cascade Mountains from 1990–1999 (refs 1, 2, 18–21). During each breeding event we quantified the number of embryos deposited, the percentage of embryonic mortality associated with *S. ferax* infections and the water depth in which embryos developed<sup>18</sup>.

Several studies have found a correlation between the summer SOI and conditions in the Pacific Northwest the following winter<sup>22</sup>. During the years in which we monitored embryo mortality we compared the relationship between summer SOI and winter precipitation in the north Cascade Mountains of Oregon. Data on the SOI were obtained from the Lamont Doherty Earth Observatory web site (<http://ingrid.ldgo.columbia.edu/>). We used standardized SOI values for the months June–November. Precipitation data was obtained from the Oregon Climate Center ([http://www.ocs.orst.edu/ocs\\_data.html](http://www.ocs.orst.edu/ocs_data.html)). We used the monthly average of all weather stations within Oregon's north Cascade Mountains (Oregon Climate Zone number 4 (ref. 22) for the months October–March).

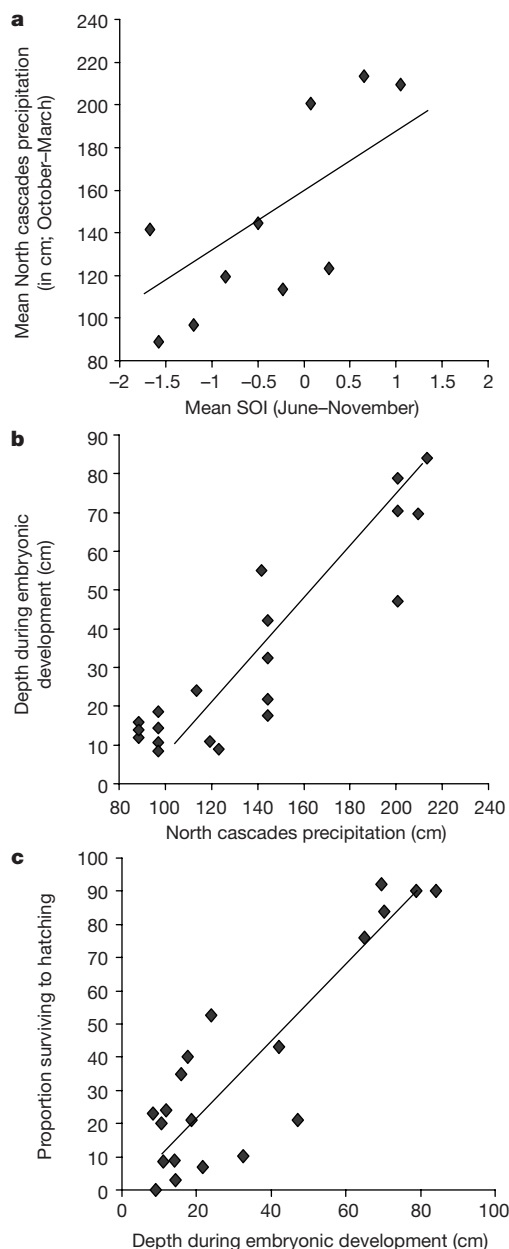
We used a field experiment to evaluate the interactive impacts of variation in water depth and UV-B exposure on *S. ferax*-associated mortality of *B. boreas* embryos. The experiment was conducted at a natural oviposition site of *B. boreas*. We manipulated the depth at which embryos were raised (10, 50 or 100 cm) and their exposure to ambient UV-B radiation (exposed to the full complement of ambient UV-B or shielded from UV-B) using a fully factorial design. We also monitored the UV-B exposure ( $\mu\text{W cm}^{-2}$ ) of embryos used in the experiment.

Observations of embryonic mortality patterns were consistent with the hypothesis that climate-induced fluctuations in the water depth influence the exposure of developing embryos to UV-B radiation. The percentage of mortality associated with *S. ferax* infection was dependent on the water depth in which embryos developed ( $R^2_{1,18} = 0.733$ ,  $P < 0.0001$ ; Fig. 1). More than 50% of the western toad embryos that developed in relatively shallow water ( $\leq 20$  cm) consistently developed *S. ferax* infections (Fig. 1). However, when eggs developed in water deeper than 45 cm, *S. ferax*-associated mortality was never more than 19% (Fig. 1). Water depth at the oviposition sites was in turn related to the amount of winter precipitation ( $R^2_{1,18} = 0.816$ ,  $P < 0.0001$ ; Fig. 1). The amount of



winter precipitation in Oregon's north Cascade Mountains during 1990–1999 was itself a function of the SOI ( $R^2_{1,8} = 0.641$ ,  $P = 0.046$ ; Fig. 1).

Results from the field experiment suggest that variation in water level influenced *S. ferax*-associated mortality patterns through exposure to UV-B radiation. The hatching success of *B. boreas* in shallow water (10 cm) was significantly affected by exposure to UV-B radiation (Table 1 and Fig. 2). The hatching success of *B. boreas* exposed to UV-B radiation was 33% less than that of their counterparts that were shielded from UV-B (Tukey's honestly significant difference (HSD),  $P < 0.001$ ). In marked contrast to patterns seen in shallow water, we found no impact of UV-B exposure on *S. ferax*-associated mortality of embryos in deep water regimes (Tukey's HSD,  $P \geq 0.86$ ). Embryos in deep water



**Figure 1** Trends in Southern Oscillation Index (SOI), precipitation, water depth and embryo mortality. **a**, Relationship between summer (June–November) SOI and winter (October–March) precipitation in the north Cascade Mountains (1989–1999). **b**, Relationship between winter precipitation (1990–1999) and water depth at oviposition sites during embryonic development. **c**, Relationship between water depth during embryonic development and *S. ferax*-associated mortality.

**Table 1** Analysis of variance on per cent survival for *B. boreas* embryos raised at different depths and exposed to varying UV-B radiation levels

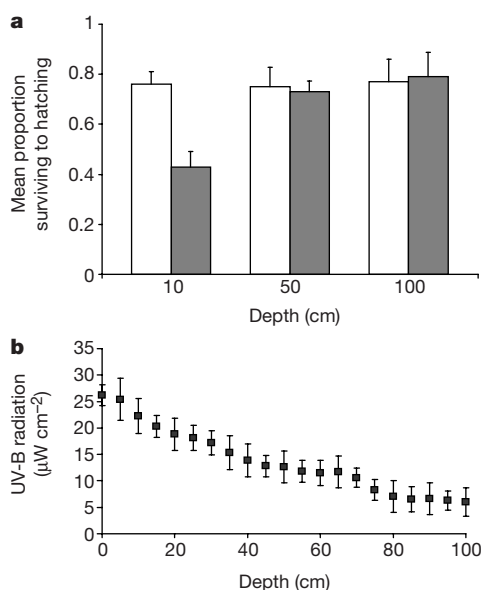
| Source of variation | d.f. | F     | P     |
|---------------------|------|-------|-------|
| Depth               | 2    | 17.78 | 0.001 |
| UV-B                | 1    | 18.73 | 0.001 |
| Depth $\times$ UV-B | 2    | 19.19 | 0.001 |
| Error               | 18   |       |       |

d.f., degrees of freedom. Error, measure of uncertainty associated with the statistical model.

(50 or 100 cm) experienced high (>75%) hatching success, regardless of UV-B exposure.

Measurements of UV-B radiation taken during the field experiment indicated dramatic differences in UV-B exposure related to water depth (Fig. 2). UV-B flux ( $\mu\text{W cm}^{-2}$ ) decreased with increasing water depth. For example, embryos raised at 50 cm received 43.5% less UV-B radiation than embryos raised in 10 cm of water.

Our findings support the hypothesis that climate-induced fluctuations in water depth have caused unusually high mortality of embryos by influencing their exposure to UV-B radiation and consequently their vulnerability to *S. ferax* infections. The changes in embryonic mortality observed for the western toad are concordant with mortality patterns for sympatric species that are also sensitive to UV-B radiation (such as the cascades frog, *Rana cascadae*)<sup>1,2,18,20,21</sup>. Disease outbreaks for these species have occurred simultaneously, suggesting that they may be related to regional changes. It has been suggested that El Niño events may be increasing in frequency and intensity as a result of global climate change<sup>7,8,23,24</sup>. Increased frequency of El Niño events may also increase the incidence of high embryonic mortality experienced by certain amphibians in the Pacific Northwest. If bouts of high embryo mortality occur with greater regularity and intensity, they may result in population declines. Although other studies have linked climate change with declines in amphibian populations<sup>4</sup>, they have not revealed specific biotic and/or physical mechanisms that underlie the declines. Our results suggest a complex series of interactions involving changes in both physical (water depth and UV-B exposure) and biotic (disease outbreaks) factors that alter mortality patterns. Because the survival of amphibians is linked closely to



**Figure 2** Field experiment results. **a**, Effects of water depth and exposure to UV-B radiation on mean hatching success ( $\pm$  s.e.m.) for *B. boreas*. Open column, ultraviolet-blocking filter; filled column, ultraviolet-transmitting filter. **b**, Measurements of UV-B flux ( $\mu\text{W cm}^{-2}$ ) at different depths.

water availability, climate changes that alter hydrology may be the precursor for similar mortality events that are believed to contribute to other population declines, including those that have been attributed to disease outbreaks<sup>3,25,26</sup>. Our results are concordant with other studies<sup>4,9</sup> that point to Pacific warming over recent decades as a common denominator for amphibian declines. The local manifestations of large-scale climate change, as well as their effects on living systems, are varied. Thus, the manner in which global climate change ultimately results in amphibian declines will probably differ among environments and species. The recurring theme of epidemic disease associated with many amphibian declines<sup>3,4,9,25,26</sup> and with declines of a wide range of other organisms<sup>27</sup>, may be explained by the impacts of climate change on lethal diseases<sup>27</sup>.

As with other studies that have examined the biological consequences of global climate fluctuations<sup>28–30</sup>, our study emphasizes the overlooked importance of ecological responses to climatic fluctuations mediated through complex local interactions. It has become increasingly clear that to predict how climate change may translate into species losses, we must link global and local processes. The challenge will be to determine whether there are general trends that may help guide these efforts. Our results agree with other work on the exposure of aquatic organisms to UV-B radiation<sup>14,15</sup> and may help to provide a general understanding on the impact of climate change on aquatic systems. These results indicate that in aquatic ecosystems, climatic changes can increase the exposure of organisms to UV-B<sup>14,15</sup>. In fact, our results suggest that in high mountain lakes, the fluctuation in water levels caused by global change could be of more concern with respect to UV-B exposure than depletion of stratospheric ozone. It is clear that the interaction of climatic changes and increased UV-B exposure in aquatic systems deserves further study. □

## Methods

### Measurements of mortality at oviposition sites

At each site we estimated the total number of eggs laid by counting the number of breeding pairs and multiplying the total number by 12,000 (the average number of eggs laid per female per breeding period)<sup>18</sup>. The water depth that embryos developed in was assessed at each breeding site by measuring the distance from the top of the egg mass to the surface of the water with a metre stick. We took measurements at each site on at least four separate dates during embryonic development. To standardize depth estimates, we took measurements from the centre of each communal egg mass and also from the north, south, east and west sides of each communal mass. These measurements were then averaged across each of the four dates.

The presence of *S. ferax* infection within *B. boreas* embryos is readily observable<sup>18,19</sup>. Infected eggs become covered with a visible crown of white hyphal filaments, and do not hatch<sup>1,18–20</sup>. The per cent mortality of eggs at each site was estimated by placing a 1-m<sup>2</sup> grid, containing squares with an area of 0.1 m<sup>2</sup>, over the egg masses. We counted the total number of dead and healthy eggs in each square. The percentage of egg mortality was averaged for each square to get an estimate for each grid. The grid was randomly assigned to five locations in the breeding area for an estimate of per cent mortality for that site. In 1998 and 1999, we assessed the per cent mortality of eggs at each site by randomly removing 100 eggs from the egg masses and counting the total number of dead and healthy eggs. This process was repeated 50 times at each site, and each measurement was averaged to give an estimate of per cent mortality for that site.

### Field experiment

The field experiment was conducted at a natural oviposition site of *B. boreas* (Lost Lake, Linn County, Oregon, 97 km east of Albany, Oregon) from 15 May to 29 May 1996. One hundred newly deposited embryos (<24 h old), ten each from ten separate clutches, were placed in each enclosure. Enclosures (27 cm × 16 cm × 11.5 cm) were composed of opaque plastic frames with floors of 500-μm fiberglass mesh screen. The mesh screen prevented the eggs from moving in or out, but allowed water flow and *S. ferax* transmission. A UV-B blocking filter made of mylar was placed over one half of the enclosures. The remaining enclosures were covered with an acetate filter that transmitted UV-B. Analyses with an Optronics International model 752 spectroradiometer showed that mylar blocked 100% of UV-B radiation. The acetate allowed about 80% UV-B transmission. Pairs of enclosures, one with a mylar filter and one with an acetate filter, were bolted on to a plastic moulded pole, and placed at one of three possible depths (10, 50 or 100 cm) below the water surface. A Hobo temperature data logger was placed with nine of the enclosure pairs, three at 10 cm, three at 50 cm and three at 100 cm, and set to record temperature six times an hour for the duration of the experiment. There were no significant differences in mean temperatures between treatments (analysis of variance (ANOVA),  $F_{2,6} = 2.23$ ,  $P = 0.189$ ).

The experiment was terminated when all of the original embryos either hatched or died. Survival was measured as the proportion of hatchlings produced per enclosure.

During the experiment UV-B measurements ( $\mu\text{W cm}^{-2}$ ) were taken with a Solar Light PMA2100 ultraviolet meter. We first measured ambient UV-B levels and then UV-B levels were measured at 5-cm intervals starting at the water surface and continuing to 100 cm below the surface. At each depth, we allowed the UV-B readings to stabilize before we moved the probe to the next depth. We repeated this process three times. Measurements were taken on three days during the experiment. All measurements were taken between 12:00 and 14:00.

### Statistical analysis

We used linear regression to test for the following effects: water depth on hatching survival; winter precipitation on water depth during embryonic development; and SOI on winter precipitation. We tested for differences in hatching success between treatments using a two-way ANOVA, with the main effects of depth (10, 50 or 100 cm) and UV-B exposure. Tukey's HSD test was used to compare treatment means where significant ( $P < 0.05$ ) differences were found with the ANOVA. We arc-sin-transformed the data on survivorship to hatching before the analysis, and after transformation the data met the parametric assumptions (normality and homogeneity of variance).

Received 22 September 2000; accepted 22 January 2001.

- Kiesecker, J. M. & Blaustein, A. R. Synergism between UV-B radiation and a pathogen magnifies amphibian embryo mortality in nature. *Proc. Natl Acad. Sci. USA* **92**, 11049–11052 (1995).
- Blaustein, A. R. et al. UV repair and resistance to solar UV-B in amphibian eggs: a link to population declines? *Proc. Natl Acad. Sci. USA* **91**, 1791–1795 (1994).
- Berger, L. et al. Chytridiomycosis causes amphibian mortality associated with population declines in the rainforests of Australia and Central America. *Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
- Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. Biological response to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
- Beebe, T. J. C. Amphibian breeding and climate. *Nature* **374**, 219–220 (1995).
- Houlahan, J. E., Findlay, C. S., Schmidt, B. R., Meyer, A. H. & Kuzmin, S. L. Quantitative evidence for global amphibian population declines. *Nature* **404**, 752–755 (2000).
- Guilderson, T. P. & Schrag, D. P. Abrupt shift in subsurface temperatures in the Tropical Pacific associated with changes in El Niño. *Science* **281**, 240–243 (1998).
- Graham, N. E. Simulation of recent global temperature trends. *Science* **267**, 661–671 (1995).
- Pounds, J. A. & Crump, M. L. Amphibian declines and climate disturbance: the case of the golden toad and the harlequin frog. *Conserv. Biol.* **8**, 72–85 (1994).
- Blumthaler, M. & Ambach, W. Indication of increasing solar ultraviolet-b radiation flux in alpine regions. *Science* **248**, 206–208 (1990).
- Kerr, J. B. & McElroy, C. J. Evidence for large upward trends of ultraviolet-b radiation linked to ozone depletion. *Science* **262**, 1032–1034 (1993).
- Stolarski, R. et al. Measured trends in stratospheric ozone. *Science* **256**, 342–349 (1992).
- Häder, D. P. Impact of UVB on aquatic organisms. *Photochem. Photobiol.* **69**, S23–S32 (1999).
- Schindler, D. W., Curtis, P. J., Parker, B. R. & Stainton, M. P. Consequences of climate warming and lake acidification for UV-B penetration in North American boreal lakes. *Nature* **379**, 705–708 (1996).
- Yan, N. D., Keller, W., Scully, N. M., Lean, D. R. S. & Dillon, P. J. Increased UV-B penetration in a lake owing to drought-induced acidification. *Nature* **381**, 141–143 (1996).
- Nussbaum, R. A., Brodie, E. D. & Storm, R. M. *Amphibian and Reptiles of the Pacific Northwest* (Idaho Univ. Press, Moscow, Idaho, 1983).
- Stebbins, R. C. & Cohen, N. W. *A Natural History of Amphibians* (Princeton Univ. Press, New Jersey, 1995).
- Kiesecker, J. M. & Blaustein, A. R. Influences of egg laying behavior on pathogenic infection of amphibian eggs. *Conserv. Biol.* **12**, 214–220 (1997).
- Blaustein, A. R., Hokit, D. G., O'Hara, R. K. & Holt, R. A. Pathogenic fungus contributes to amphibian losses in the Pacific Northwest. *Biol. Conserv.* **67**, 251–254 (1994).
- Kiesecker, J. M. & Blaustein, A. R. Pathogen reverses competition between larval amphibians. *Ecology* **80**, 2442–2448 (1999).
- Blaustein, A. R. & Kiesecker, J. M. in *The Effects of Ozone Depletion on Aquatic Ecosystems* (ed. Häder, D. P.) 175–188 (R. G. Landes, Austin, Texas, 1997).
- Redmond, K. T. & Koch, R. W. Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indexes. *Water Resour. Res.* **27**, 2381–2399 (1991).
- Meehl, G. A. & Washington, W. M. El Niño-like climate change in a model with increased atmospheric CO<sub>2</sub> concentrations. *Nature* **382**, 56–60 (1996).
- Timmermann, A. et al. Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature* **398**, 694–697 (1999).
- Lips, K. R. Decline of a tropical montane amphibian fauna. *Conserv. Biol.* **12**, 106–117 (1996).
- Laurance, W. F., McDonald, K. R. & Speare, R. Epidemic disease and the catastrophic declines of Australian rain forest frogs. *Conserv. Biol.* **10**, 406–413 (1996).
- Epstein, P. R. Perspectives: Medicine, climate and health. *Science* **285**, 347–348 (1999).
- Hughes, L. Biological consequences of global warming: is the signal already apparent? *Trends Ecol. Evol.* **15**, 56–61 (2000).
- Post, E., Peterson, R. O., Stenseth, N. C. & McKaren, B. E. Ecosystem consequences of wolf behavioural response to climate. *Nature* **401**, 905–907 (1999).
- Lima, M., Keymer, J. E. & Jaksic, F. M. El Niño-Southern oscillation-driven rainfall variability and delayed density dependence cause rodent outbreaks in western South America: linking demography and population dynamics. *Am. Nat.* **153**, 476–491 (1999).

### Acknowledgements

We thank the U.S. Forest Service, Pacific Northwest Research Station and D. H. Olson for the use of vehicles and temperature data loggers, and D. G. Hokit and S. Walls for

discussions regarding experimental design. We also thank G. Taylor from the Oregon State Climate Center for help with obtaining the climate data. J. A. Pounds, E. Post, M. Vronsky and N. Chevoteravich provided helpful suggestions on this manuscript. Funding was provided by the Declining Amphibian Population Task Force Seed Grant (to J.M.K.); NIH/NSF Ecology of Infectious Diseases grant (to J.M.K.); and the Department of Biology, Pennsylvania State University.

Correspondence and requests for materials should be addressed to J.M.K. (e-mail: jmk23@psu.edu).

## Towards a resolution of the lek paradox

Janne S. Kotiaho\*, Leigh W. Simmons & Joseph L. Tomkins\*

Evolutionary Biology Research Group, Department of Zoology, University of Western Australia, Nedlands, Western Australia 6907, Australia

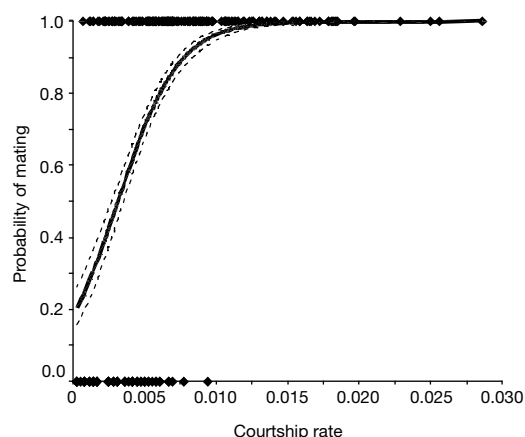
Genetic benefits in the shape of 'good genes' have been invoked to explain costly female choice in the absence of direct fitness benefits<sup>1–3</sup>. Little genetic variance in fitness traits is expected, however, because directional selection tends to drive beneficial alleles to fixation<sup>4–6</sup>. There seems to be little potential, therefore, for female choice to result in genetic benefits, giving rise to the 'lek paradox'<sup>7–9</sup>. Nevertheless, evidence shows that genetic variance persists despite directional selection<sup>10,11</sup> and genetic benefits of female choice are frequently reported<sup>12,13</sup>. A theoretical solution to the lek paradox has been proposed on the basis of two assumptions<sup>14</sup>: that traits are condition-dependent, and that condition shows high genetic variance. The observed genetic variability in sexual traits will be accounted for, because a proportion of the genetic variance in condition will be captured and expressed in the trait<sup>14</sup>. Here we report results from experiments showing that male courtship rate in the dung beetle *Onthophagus taurus* is a condition-dependent trait that is preferred by females. More importantly, male condition has high genetic variance and is genetically correlated with courtship rate. Our results thereby represent a significant step towards a resolution of the lek paradox.

High genetic variance in condition and condition-dependent expression of sexual traits are two critical assumptions underlying the condition capture model for the evolution of costly female preferences in the absence of direct benefits<sup>1,14</sup>. When these predictions are empirically validated, the long-standing problem of sexual selection of good genes, the lek paradox, will be resolved<sup>7–9,14</sup>. Although there is abundant evidence for condition-dependent expression of traits<sup>3,15–18</sup>, direct empirical evidence for genetic variance in condition is rare<sup>19–21</sup>.

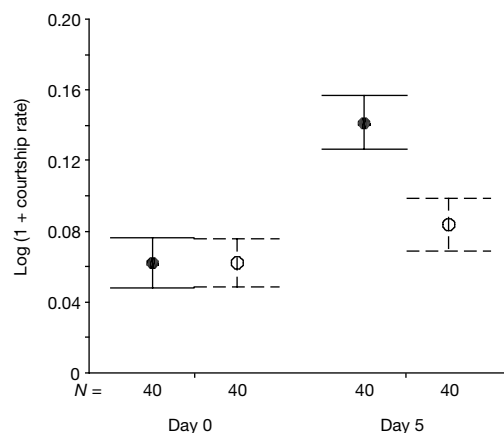
We used random mating trials to study female preference for male courtship rate in the dung beetle *O. taurus*. Males used in this experiment were derived from a half-sib breeding design<sup>5,6</sup>, and females were laboratory-reared virgins selected at random from a large culture population. Male dung beetles cannot copulate unless they can persuade the female to open her genital tergite. To do this, a male will court a female by tapping her back with his head and forelegs in bouts lasting a few seconds. We excluded the possibility of male competition affecting our results by introducing only one male and one female into an artificial dung beetle tunnel. Courtship rate was calculated as the number of courtship bouts per unit time. We observed 232 pairs, of which 170 males were successful in

mating. The probability of mating was strongly affected by the male's courtship rate (logistic regression  $\chi^2_1 = 79.76$ ,  $n = 232$ ,  $P < 0.0001$ ; effect size estimated as Pearson's correlation coefficient  $r = 0.6$ ; Fig. 1). To ensure that this result was robust and not biased by measures derived from related males, we replicated the experiment using males collected from a field population: this replication confirmed the above result (logistic regression  $\chi^2_1 = 43.30$ ,  $n = 80$ ,  $P < 0.0001$ ; effect size  $r = 0.7$ ). The preference function shown in Fig. 1 resulted in moderate directional selection for higher courtship rate (intensity of directional selection ( $i$ ) with 95% confidence intervals (CI) and significance test:  $i = 0.331$ ,  $CI_{lower} = 0.121$ ,  $CI_{upper} = 0.491$ ;  $t_{400} = 3.20$ ,  $P = 0.0015$ ; methods as described<sup>22,23</sup>).

We experimentally examined the condition-dependence of courtship rate by manipulating the condition of males through their nutritional state. We measured the initial courtship rate and body mass of 80 field-collected males and randomly allocated them to one of two food treatments (constant supply of food or no food). The mean and variance in initial courtship rate or body mass did not differ between the groups (means: analysis of variance (ANOVA)  $F_{1,78} = 0.00$ ,  $P = 0.9876$  and  $F_{1,78} = 0.00$ ,  $P = 0.9823$ , respectively; variances: Levene's test,  $F_{1,78} = 0.14$ ,  $P = 0.7065$  and  $F_{1,78} = 0.05$ ,  $P = 0.8302$ , respectively). After five days of food manipulation, the courtship rates and body masses were remeasured. The manipulation had a substantial effect on courtship rate: males with



**Figure 1** Probability of mating  $\pm$  s.e. as a function of courtship rate. Diamonds, original data. The function was estimated using cubic spline nonparametric regression with FORTRAN77 computer routines<sup>23</sup>. Standard errors were derived from 1,000 bootstrap replications.



**Figure 2** Condition dependence of courtship rate. Left, mean  $\pm$  s.e. of courtship rate per minute (log + 1 transformed); right, the same after five days of manipulation of food availability. Solid symbols, constant food treatment; open symbols, no food treatment.

\* Present addresses: Department of Biological and Environmental Sciences, University of Jyväskylä, PO Box 35, FIN-40351 Jyväskylä, Finland (J.S.K.); Department of Environmental and Evolutionary Biology, University of St Andrews, St Andrews, Fife, Scotland, UK (J.L.T.).



**Table 1 Nested analysis of variance for courtship rate**

| Source     | SS                      | d.f. | MS                        | F    | P      | Eta <sup>2</sup> * |
|------------|-------------------------|------|---------------------------|------|--------|--------------------|
| Date†      | 1.28 × 10 <sup>-4</sup> | 1    | 1.28 × 10 <sup>-4</sup>   | 0.21 | 0.6491 | 0.002              |
| Sire       | 1.73 × 10 <sup>-2</sup> | 11   | 1.57 × 10 <sup>-3</sup> ‡ | 2.16 | 0.0495 | 0.462              |
| Dam (Sire) | 1.77 × 10 <sup>-2</sup> | 24   | 7.38 × 10 <sup>-4</sup>   | 1.20 | 0.2517 | 0.178              |
| Error      | 8.17 × 10 <sup>-2</sup> | 133  | 6.14 × 10 <sup>-4</sup>   |      |        |                    |

Courtship rate is square-root transformed. d.f., degrees of freedom. SS, sums of squares; MS, mean square; F, test statistics; P, probability.

\* Eta<sup>2</sup> refers to the proportion of variance explained.

† Courtship rate tended to covary with the date of the observation. This is accounted for by using the date of observation as a covariate.

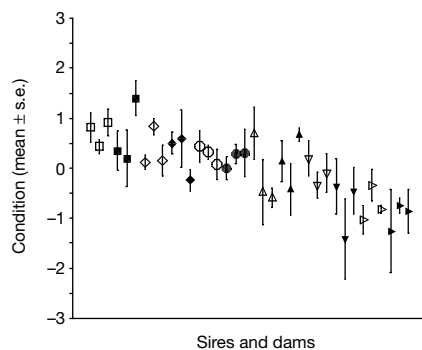
‡ To account for unequal sample sizes of offspring within sires, the error term for sires was calculated using Satterthwaite's approximation: 0.918 MS(Dam(Sire)) + 0.082 MS(Error) = 7.28 × 10<sup>-4</sup>.

constant food had significantly higher courtship rates than males with no food, indicating that, as predicted, courtship rate covaries positively with condition (repeated measures ANOVA  $F_{1,78} = 6.15$ ,  $P = 0.0153$ ; Fig. 2).

In addition to courtship rate, male body mass changed depending on the food treatment. Males with constant food increased in mass, whereas males with no food decreased in mass (repeated measures ANOVA  $F_{1,78} = 160.37$ ,  $P < 0.0001$ ). As there was no difference in the initial body mass between the treatments, our manipulation was effective in increasing and decreasing the body condition. Moreover, we estimated the relative condition of males within the two treatments as residual mass from linear regression between log (body mass) and log (pronotum width)<sup>24,25</sup>, and found that the courtship rate of males with no food covaried positively with condition ( $r = 0.35$ ,  $n = 40$ ,  $P = 0.0268$ ) whereas there was no such covariation with constant food ( $r = -0.09$ ,  $n = 40$ ,  $P = 0.5808$ ). These two correlations are marginally different from each other ( $Z = 1.93$ ,  $P = 0.0536$ ). The difference is expected because sexual selection theory predicts that only individuals of high genetic quality are able to bear the costs of trait expression under harsh conditions, whereas under good conditions the costs of expressing a trait decrease, thereby reducing the difference in trait expression between males<sup>3</sup>.

We estimated the genetic variance in courtship rate and condition using a standard half-sib breeding design<sup>5,6</sup>. The data were analysed with nested analyses of variance, having sire and dam as random factors with the latter nested within the former. The genetic correlation between courtship rate and condition was estimated from nested multivariate analysis of variance.

There was a significant sire effect, indicating that there is genetic variance in male courtship rate (Table 1). The heritability estimate  $h^2_{\text{sire}}$  (2 standard errors (s.e.))<sup>6</sup> of courtship rate was moderate 0.337 (0.288) and corresponds well with heritability estimates generally reported for sexual traits under directional selection<sup>11</sup>. What makes our results novel, however, is that we found a strong sire effect on condition itself (Table 2; Fig. 3). Our result shows that 67% of total



**Figure 3** Genetic variation in condition. Condition is measured as the standardized residual mass from a linear regression of log body mass on log pronotum width. The different symbols represent the 12 sires. The three replications within each sire cluster represent the three dams for each sire. Sires are ranked according to the mean condition of their offspring.

**Table 2 Nested analysis of variance for condition**

| Source     | SS    | d.f. | MS    | F    | P      | Eta <sup>2</sup> * |
|------------|-------|------|-------|------|--------|--------------------|
| Date†      | 6.01  | 1    | 6.01  | 8.37 | 0.0045 | 0.059              |
| Sire       | 52.03 | 11   | 4.73‡ | 5.43 | 0.0001 | 0.684              |
| Dam (Sire) | 21.22 | 24   | 0.88  | 1.23 | 0.2269 | 0.182              |
| Error      | 95.50 | 133  | 0.72  |      |        |                    |

\* Eta<sup>2</sup> refers to the proportion of variance explained.

† Condition tended to covary with the date of the observation. This is accounted for by using the date of observation as a covariate.

‡ To account for unequal sample sizes of offspring within sires, the error term for sires was calculated using Satterthwaite's approximation: 0.918 MS(Dam(Sire)) + 0.082 MS(Error) = 0.87.

variation in condition was genetic in origin. Furthermore, the estimate of the genetic correlation  $r_{\text{Asire}}$  (2 s.e.)<sup>6</sup> between male condition and courtship rate was positive and relatively high (0.709 (0.263)). To assess the levels of additive genetic variance ( $V_A$ ) in courtship rate and in condition, we calculated the coefficient of variation for the additive genetic variances ( $CV_A$ ) and compared these with published values of  $CV_A$  for sexually selected traits<sup>11</sup> and for two fitness components, fecundity and longevity<sup>10</sup>. The median value of  $CV_A$  for sexually selected traits was 8.2 (derived from Table 2 in ref. 11), and for fecundity and longevity 11.9 and 9.9, respectively<sup>10</sup>. Our estimate of  $CV_A$  (2 s.e.) for courtship rate was 8.57 (0.94) and for condition 27.05 (3.14).  $CV_A$  for courtship rate corresponds well with values previously reported for sexually selected traits. However, the  $CV_A$  for condition was more than twice those reported for fitness components. This comparison indicates that the level of additive genetic variance for condition is high, as predicted<sup>14</sup>.

Our comprehensive set of experimental tests provides empirical support for the condition capture model proposed to explain the maintenance of costly female choice on the basis of genetic benefits<sup>1,14</sup>. As such, our study represents a step towards the resolution of the lek paradox. The condition capture model also predicts that over evolutionary time, the amount of genetic variation for sexual traits should increase as they become increasingly condition-dependent. Such an evolutionary association has recently been reported across species of stalk-eyed flies<sup>26</sup>. Ultimately, the theory relies on the assumption that condition is influenced by a large number of loci, resulting in a relatively high frequency of mutations in genes coding for condition<sup>14</sup>. Thus, a complete resolution of the lek paradox will require that the frequency of mutations in genes that contribute to genetic variation in condition be quantified. □

## Methods

### Female choice

Artificial dung beetle tunnels were constructed from 60-mm-long clear rectangular plastic vials measuring 13 mm in width and 36 mm in depth. Vials were half filled with plaster of Paris to create a 60 mm long, 13 mm wide and 17 mm high tunnel. Tunnels were smeared with cow dung and dried. Before trials, tunnels were moistened with water. One randomly selected female and male were introduced into a tunnel and the courtship bouts of the males were observed constantly until mating occurred or for 60 min. No individual was used more than once. Before courtship and female choice observations, male offspring were housed individually in identical small sand-filled containers (7 × 7 × 5 cm<sup>3</sup>) with a constant supply of fresh dung.

### Condition-dependence

At the start of the experiment we measured the initial body mass (to the nearest 0.01 mg), pronotum width (to the nearest 0.05 mm) and courtship rate of 80 males. Courtship rate was observed as described above. As some males mated during the observations, the time for courtship observations varies. Therefore, courtship rate per minute rather than total number of courtship bouts was used in the analysis.

After initial measurements, males were divided between two food ration treatments, with the two groups matched for mean and variance in morphological measures and initial courtship rates (see text). After measurements, beetles were housed individually in small plastic containers (7 × 7 × 5 cm<sup>3</sup>) two-thirds filled with moist sand. Constant food ration males were provided with a supply of fresh dung and no food ration males were provided with moist sand only. After five days of food manipulation, the measurements of male body mass and courtship rate were repeated as described above.

## What is condition?

Rowe and Houle<sup>14</sup> referred to condition as the pool of resources available for utilization. This definition of condition corresponds to residual reproductive value or state in life history models and accounts for a large proportion of fitness<sup>14</sup>. This definition means that condition cannot be measured directly. Therefore, we used the residual mass as an estimate of condition<sup>24,25</sup>. Intuitively, residual mass approximates well the resources that should be available for utilization and there is likely to be a relationship between the two, but we acknowledge that this is not exactly the same as condition as used by Rowe and Houle<sup>14</sup>.

## Genetic variances and correlation

To estimate the genetic variance in male courtship rate and in male condition we used a standard half-sib breeding design<sup>5,6</sup>. We housed each of the 12 field-collected sires with three randomly selected F<sub>1</sub> laboratory-reared virgin dams. Dams did not differ in body size across sires. After five days, dams were set up individually to construct brood masses. As males of this species occasionally help in constructing the brood mass, non-genetic paternal effects may occur. This was taken into consideration in the experimental design by excluding males after mating so that females constructed brood masses alone. Thus, our experimental design excluded the possibility of non-genetic paternal effects.

It has also been suggested that females may differentially invest in their offspring depending on the attractiveness of the male<sup>27–29</sup>. Differential investment based on male attractiveness could result in non-genetic similarities between the offspring of dams allocated to the same sire. If there is differential investment, apparent genetic sire effects may be confounded by non-genetic maternal effects. This problem of interpretation is inherent in all half-sib breeding designs where female investment has not been measured. In our experiments we took this possibility into account by measuring the female investment in offspring. There is variation in the weight of brood masses constructed by females, which is known to have a strong effect on offspring size and development. Thus, we measured the female investment as the brood mass weight. We found evidence for differential investment in brood mass weight translating into differential investment in offspring size (unpublished data), but brood mass weight did not influence offspring condition ( $r = -0.042$ ,  $n = 167$ ,  $P = 0.596$ ) or courtship rate ( $r = -0.053$ ,  $n = 228$ ,  $P = 0.427$ ). Therefore, our results do not arise from differential investment by females. However, we did not measure the size of the egg itself in the brood mass. Nevertheless, offspring growth and development in dung beetles is almost entirely dependent on maternal investment in brood mass<sup>30</sup> so that relatively subtle differences in the allocation of resources into eggs, if they existed, are unlikely to contribute to the sire effects reported here.

Received 9 October 2000; accepted 16 January 2001.

- Andersson, M. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. *Evolution* **40**, 804–816 (1986).
- Pomiankowski, A. N. The evolution of female mate preferences for male genetic quality. *Oxford Surv. Evol. Biol.* **5**, 136–184 (1988).
- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, 1958).
- Falconer, D. S. *Introduction to Quantitative Genetics* (Longman, New York, 1989).
- Roff, D. *Evolutionary Quantitative Genetics* (Chapman & Hall, New York, 1997).
- Borgia, G. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 19–80 (Academic, New York, 1979).
- Taylor, P. D. & Williams, G. C. The lek paradox is not resolved. *Theor. Pop. Biol.* **22**, 392–409 (1982).
- Kirkpatrick, M. & Ryan, M. J. The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38 (1991).
- Houle, D. Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195–204 (1992).
- Pomiankowski, A. & Møller, A. P. A resolution of the lek paradox. *Proc. R. Soc. Lond. B* **260**, 21–29 (1995).
- Møller, A. P. & Alatalo, R. V. Good-genes effects in sexual selection. *Proc. R. Soc. Lond. B* **266**, 85–91 (1999).
- Alatalo, R. V., Kotiaho, J., Mappes, J. & Parri, S. Mate choice for offspring performance: major benefits or minor costs? *Proc. R. Soc. Lond. B* **265**, 2297–2301 (1998).
- Rowe, L. & Houle, D. The lek paradox and the capture of genetic variance by condition dependent traits. *Proc. R. Soc. Lond. B* **263**, 1415–1421 (1996).
- Johnstone, R. A. Sexual selection, honest advertisement and the handicap principle: reviewing the evidence. *Biol. Rev.* **70**, 1–65 (1995).
- Kotiaho, J. S. Testing the assumptions of conditional handicap theory: costs and condition dependence of a sexually selected trait. *Behav. Ecol. Sociobiol.* **48**, 188–194 (2000).
- Griffith, S. C., Owens, I. P. F. & Burke, T. Environmental determination of a sexually selected trait. *Nature* **400**, 358–360 (1999).
- David, P., Bjorksten, T., Fowler, K. & Pomiankowski, A. Condition-dependent signalling of genetic variation in stalk-eyed flies. *Nature* **406**, 186–188 (2000).
- Merilä, J. Genetic variation in offspring condition: an experiment. *Funct. Ecol.* **10**, 465–474 (1996).
- Sheldon, B. C., Merilä, J., Quarnström, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution of offspring condition predicted by size of male secondary sexual character. *Proc. R. Soc. Lond. B* **264**, 297–302 (1997).
- Merilä, J., Przybylo, R. & Sheldon, B. C. Genetic variation and natural selection on blue tit body condition in different environments. *Genet. Res.* **73**, 165–176 (1999).
- Arnold, S. J. & Wade, M. J. On the measurement of natural and sexual selection: theory. *Evolution* **38**, 709–719 (1984).
- Schluter, D. Estimating the form of natural selection on a quantitative trait. *Evolution* **42**, 849–861 (1988).

- Jakob, E. M., Marshall, S. D. & Uetz, G. W. Estimating fitness: a comparison of body condition indices. *Oikos* **77**, 61–67 (1996).
- Kotiaho, J. S. Estimating fitness: comparison of body condition indices revisited. *Oikos* **87**, 399–400 (1999).
- Wilkinson, G. S. & Taper, M. Evolution of genetic variation for condition-dependent traits in stalk-eyed flies. *Proc. R. Soc. Lond. B* **266**, 1685–1690 (1999).
- Gil, D., Graves, J., Hazon, N. & Wells, A. Male attractiveness and differential testosterone investment in zebra finch eggs. *Science* **286**, 126–128 (1999).
- Cunningham, E. J. & Russell, A. F. Egg investment is influenced by male attractiveness in the mallard. *Nature* **404**, 74–77 (2000).
- Sheldon, B. Differential allocation: tests, mechanisms and implications. *Trends Ecol. Evol.* **15**, 397–402 (2000).
- Hunt, J. & Simmons, L. W. Patterns of fluctuating asymmetry in beetle horns: an experimental examination of the honest signalling hypothesis. *Behav. Ecol. Sociobiol.* **41**, 109–114 (1997).

## Acknowledgements

We thank J. Hunt for help during breeding of the beetles; R. Alatalo and J. Reynolds for discussions; and R. Alatalo for comments on the manuscript. J.S.K. was supported by the Academy of Finland, L.W.S. by the ARC and J.L.T. by a postdoctoral research fellowship from the University of Western Australia.

Correspondence and request for materials should be addressed to J.S.K. (e-mail: jkotiaho@cc.jyu.fi).

# Hyperacute directional hearing in a microscale auditory system

Andrew C. Mason\*, Michael L. Oshinsky† & Ron R. Hoy

Neurobiology and Behaviour, SG Mudd Hall, Cornell University, Ithaca, New York 14853, USA

The physics of sound propagation imposes fundamental constraints on sound localization: for a given frequency, the smaller the receiver, the smaller the available cues<sup>1</sup>. Thus, the creation of nanoscale acoustic microphones with directional sensitivity is very difficult. The fly *Ormia ochracea* possesses an unusual 'ear' that largely overcomes these physical constraints<sup>2–5</sup>; attempts to exploit principles derived from *O. ochracea* for improved hearing aids are now in progress<sup>6</sup>. Here we report that *O. ochracea* can behaviourally localize a salient sound source with a precision equal to that of humans<sup>7</sup>. Despite its small size and minuscule interaural cues, the fly localizes sound sources to within 2° azimuth. As the fly's eardrums are less than 0.5 mm apart, localization cues are around 50 ns. Directional information is represented in the auditory system by the relative timing of receptor responses in the two ears. Low-jitter, phasic receptor responses are pooled to achieve hyperacute timecoding<sup>8,9</sup>. These results demonstrate that nanoscale/microscale directional microphones patterned after *O. ochracea* have the potential for highly accurate directional sensitivity, independent of their size. Notably, in the fly itself this performance is dependent on a newly discovered set of specific coding strategies employed by the nervous system.

*Ormia ochracea* (Diptera: Tachinidae) is a parasitoid fly<sup>10,11</sup>. Gravid female flies locate their hosts, male crickets, by homing in on their loud, persistent songs. Because of its small body size, *O. ochracea* must deal with extremely small interaural difference cues to guide directional hearing<sup>3</sup>. The host cricket's calling song is an amplitude-modulated 5 kHz tone (6.8 cm wavelength); however, the fly measures less than 1 cm in any aspect and the distance

\* Present address: Division of Life Sciences, University of Toronto at Scarborough, 1265 Military Trail, Scarborough, Ontario M1C 1A4, Canada.

† Present address: Department of Neurology, Jefferson Headache Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA.

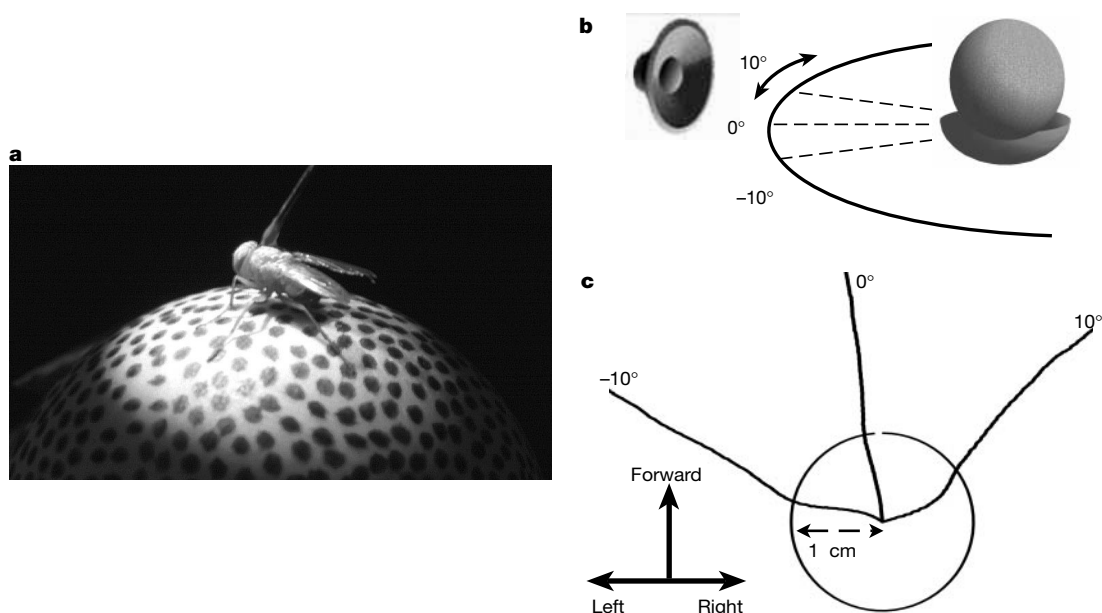
between its eardrums is about 0.5 mm. This means that 5-kHz sound waves are not diffracted by the fly's body and generate no interaural intensity difference; indeed, none can be measured<sup>3</sup>. The interaural time difference (ITD) is frequency-independent and depends only on the speed of sound and the distance between the two ears. The maximal ITD in *O. ochracea* at 90° azimuth is 1.5  $\mu$ s, and decreases to zero for a sound source on the midline axis (0° azimuth). This minuscule ITD is the only physical cue available for computation of source direction; yet *O. ochracea* can reliably localize cricket song both in nature and in the laboratory<sup>10–12</sup>. The principal evolutionary innovation responsible for the ability of *O. ochracea* to overcome its unfavourable auditory physics is a pair of anatomically and functionally coupled eardrums<sup>3,4</sup>. The mechanical resonance of *O. ochracea*'s peripheral auditory apparatus in a directional sound field transforms the minuscule time delay in the free field into two cues that can be used by its nervous system. First, the interaural time delay between the ipsilateral and contralateral eardrums is increased from a maximum of 1.5  $\mu$ s to about 55  $\mu$ s. Second, the vibration amplitudes of the ipsilateral eardrum are up to 10 dB greater than those of the contralateral, for sound sources at 45–90° azimuth. Thus, minute ITDs in the sound field are converted by the tympanal mechanics to interaural differences that are processed neurally, as described below.

To measure the auditory directional acuity of *O. ochracea*, we tethered females on a spherical treadmill (see Methods). The treadmill was attached to a moveable speaker, which could be positioned around the fly with 0.5° accuracy (Fig. 1). The flies walked in response to simulated cricket chirps broadcast from the speaker, and their movements were recorded by computer. Flies tracked changes in speaker location and altered their walking trajectory accordingly (Fig. 1b, c). Response latency was consistent (mean  $\pm$  s.d.; 46.6  $\pm$  4.9 ms), and similar to freely walking conditions (49  $\pm$  22 ms). The duration of walking responses exceeded the stimulus duration (200 ms), and in most cases flies continued to walk with a constant velocity throughout the 1.2-s recording interval. Flies reliably turned in the direction of

the speaker (Figs. 1c, 2a) and the turns were graded: larger turns in response to larger speaker angles up to 20–30° (Fig. 2b), the maximal response.

We measured the directional sensitivity of flies in two ways. First, by repeatedly presenting stimuli from speaker positions of the same angular displacement on either side of flies, we were able to measure their accuracy in lateralizing the sound source as a function of angle of incidence (that is, simply discriminating rightward from leftward speaker locations). Second, we compared the mean angular headings for different speaker locations on the same side as a measure of the flies' ability to truly localize the sound source (see Methods). Flies were extremely sensitive to changes in the location of the speaker by both measures. Overall, flies showed significant discrimination of sound angles as small as  $\pm 2^\circ$ , and some individuals were able to discriminate speaker angles of  $\pm 1^\circ$  from the midline (Fig. 2c, d). In the localization task, mean walking paths for individual flies and pooled responses of seven flies were significantly different, even for speaker locations on the same side of the midline differing by only 1–2° (Fig. 2e, f). Thus, flies are not only able to lateralize a sound source lying very close to the midline axis, but can also truly localize the source by making very precise turning movements. As the fly's two eardrums are separated by only 0.5 mm, a 1° angle of incidence represents an ITD cue at the tympanal membranes of a mere 25 ns. Similarly, a change in speaker location from 1° to 2° azimuth corresponds to a change in ITD from 25 ns to 50 ns. These minute time differences are the only directional cues available to the auditory system of the fly.

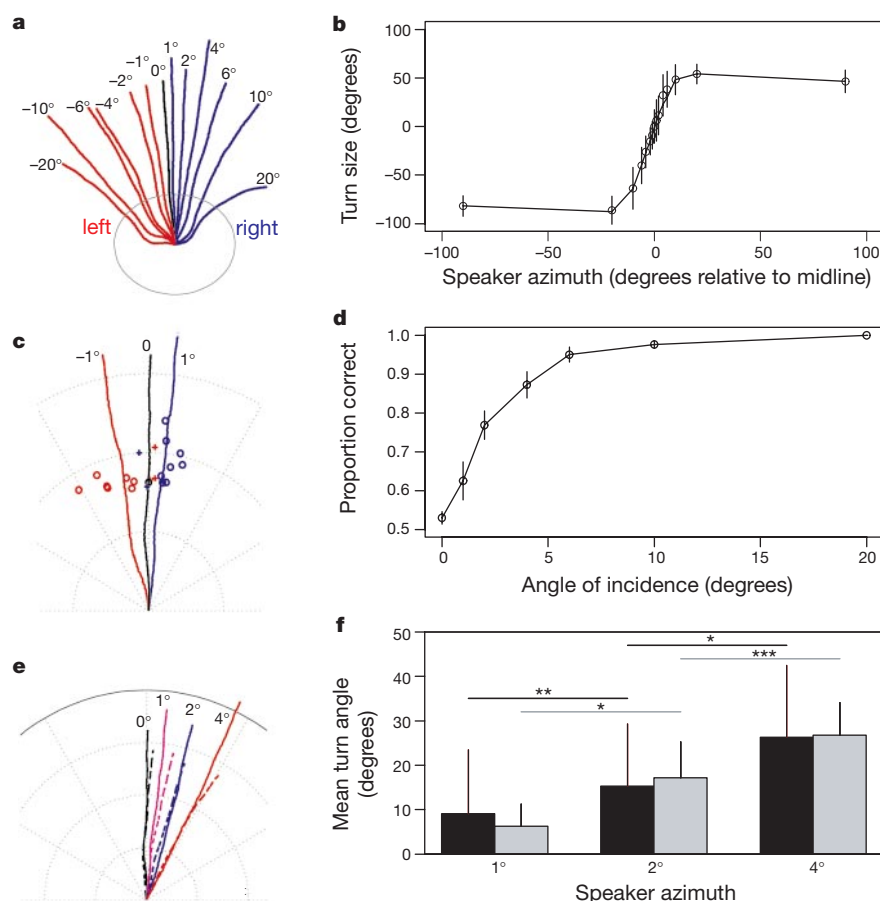
What are the neurosensory mechanisms that underlie the flies' behavioural ability to localize sound with such remarkable accuracy? We recorded the responses of individual auditory receptors. Each of the fly's ears contains about 100 receptor neurons<sup>5,13</sup>. Auditory receptors varied in their absolute sensitivity to sound level, with single unit thresholds at 5 kHz ranging from 55–95 dB sound pressure level (SPL) (mean  $\pm$  s.d.; 77.0  $\pm$  8.3 dB SPL,  $n = 28$ ). Three characteristics of receptor physiology suggest that auditory directionality is based on a spike time code at the receptor



**Figure 1** Behavioural measurement of auditory directional acuity. **a**, Female *O. ochracea* tethered on a speckled spherical treadmill. **b**, Simulated cricket calls broadcast from a moveable speaker elicit phonotactic walking responses from the fly. The resulting rotation of the ball is detected by an optical sensor and recorded by computer. **c**, Virtual walking

path of *O. ochracea* in response to stimuli from different directions. The forward direction (on the midline axis) is 0°. Negative angles represent speaker positions to the left of midline; positive angles to the right. The circle corresponds to a 1 cm radius of real walking distance.





**Figure 2** Directional responses. **a**, Mean paths of one fly ( $n = 10$  trials for each speaker position). **b**, Turn size (degrees of rotation relative to  $0^\circ$  path, mean  $\pm$  circular s.d.), measured at midpoint of each run ( $n = 7$  flies, 10 trials per fly per angle). **c**, Sound lateralization. Responses of a fly with mean paths significantly different at  $\pm 1^\circ$  ( $n = 10$ , mean path angles: left,  $-10.0 \pm 11.0^\circ$ , Rayleigh's  $Z = 0.9566$ ; right,  $11.1 \pm 4.9^\circ$ ,  $Z = 0.9962$ ; left versus zero,  $F = 7.8442$ ,  $P < 0.01$ ; right versus zero,  $F = 5.4731$ ,  $P < 0.025$ )<sup>18</sup>. Symbols indicate midpoints of individual runs. Individual runs for all flies and angles were scored (circles, correct; plus signs, incorrect) depending on whether they

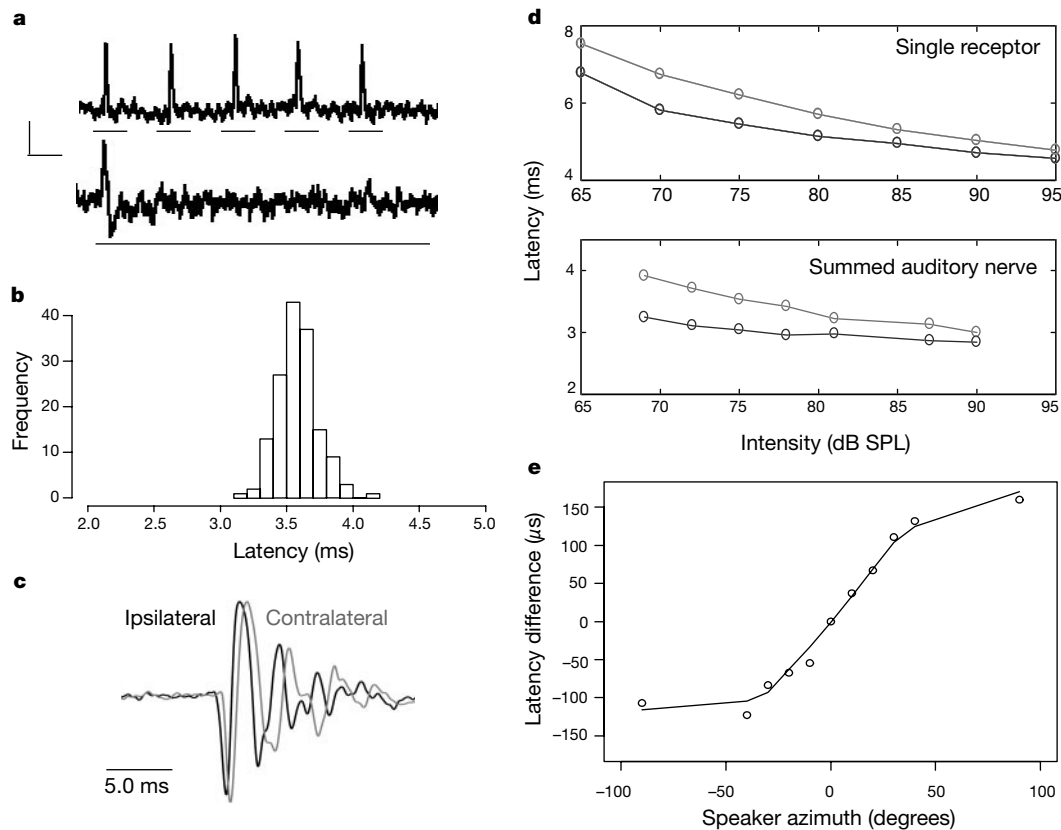
were displaced from the mean  $0^\circ$  run in the same direction as the speaker. **d**, Mean ( $\pm$  s.e.) proportion of correct turns for all speaker angles (right and left pooled,  $n = 19$  flies, 20 turns per angle per fly). **e**, Sound localization. Mean paths ( $n = 10$  runs per angle) for one fly (broken lines), and all flies pooled (solid lines,  $n = 7$ ) are shown. **f**, Turn angles (mean  $\pm$  s.d.) for individual (black) and pooled (grey) responses varied with speaker azimuth. Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.01$ ; triple asterisk,  $P < 0.001$ .

level. First, we found that most receptors responded to a sound pulse with only a single spike, independently of sound level or duration (Fig. 3a); they never produced additional spikes with increasing sound level. Second, increases in stimulus intensity were instead coded by response latency—the mean latency for a given receptor systematically decreased as stimulus intensity increased (Fig. 3d). Third, at a given sound level the response latency was very consistent, with an average jitter (s.d. of the latency) of  $70 \mu\text{s}$  (Fig. 3b). This may be compared with jitter at the millisecond level in primate cortical neurons<sup>14</sup>, but only  $30 \mu\text{s}$  for the specialized time-coding receptors in the electric fish *Eigenmannia*<sup>15</sup>. The individual auditory receptors in *O. ochracea* fire with highly consistent latency, the value of which is a function of stimulus intensity (Fig. 3d) resembling the situation in electric fish. It is important to realize that differences in sound direction at the auditory periphery cause a difference in the relative amplitude at which the eardrums vibrate<sup>3</sup>—equivalent to interaural differences in stimulus intensity. Thus, sound direction can be represented by the relative timing of receptor spikes in the left and right auditory channels.

To verify the effects of sound direction on the timing of afferent auditory responses at the population level, we made simultaneous extracellular recordings from the two auditory nerves while changing the location of a sound source. The time course of the auditory

summed action potentials (SAPs) matched the timing and threshold characteristics of individual receptors (Fig. 3b, c). Notably, the dependence of SAP latencies on stimulus intensity matched the latency/intensity relationship shown by individual receptors. In fact, overall population response latencies reflected those of individual receptors over the full range of intensities tested (Fig. 3c, d). Direct measurement of the mean interaural latency differences in auditory responses as a function of sound direction showed that they varied with angle of incidence for angles up to  $20$ – $30^\circ$  but saturated for higher angles, mirroring the fly's behavioural responses (Figs 2b, 3e). Thus, interaural latency differences can encode the position of a sound source because the time-preserving characteristics at the single-unit level confer sufficient temporal coherence to the afferent population comprising each auditory channel.

Nevertheless, difference cues at the receptor level remain extremely small. Interaural latency differences are about  $3.5 \mu\text{s}$  per degree over a range of  $30^\circ$  azimuth on either side of the midline. Therefore, for the smallest sound angles consistently discriminated by *O. ochracea* ( $2^\circ$ ), the overall latency difference is about  $7 \mu\text{s}$ . The variability of receptor-spike timing ( $70 \mu\text{s}$  jitter for individual receptors) is much greater. Hence, reliable detection of a mean latency difference of  $7 \mu\text{s}$  can only be achieved by pooling the responses of a population



**Figure 3** Characteristics of auditory receptors. **a**, Most receptors respond with only a single spike at stimulus onset, independently of duration. Responses are shown for a simulated cricket chirp (top) and continuous tone (bottom). Lines indicate stimulus timing (5 kHz, 85 dB SPL). Scales: horizontal, 10 ms; vertical, 2 mV. **b**, Spike latency for a single receptor. **c**, Summed action potentials recorded in the left (black) and right (grey) auditory nerves in response to speaker at  $-90^\circ$  (50 sweeps averaged). The effect of direction on response latency is apparent (ipsilateral response leads). **d**, Latency versus stimulus

intensity in single receptors (upper, speakers at  $\pm 90^\circ$ ) and summed action potentials (lower, speaker at  $-90^\circ$ ). Responses shown are for  $-90^\circ$  speaker location. Latency declines with increasing intensity in both ipsilateral (black) and contralateral (grey) responses, but ipsilateral latencies are shorter at all intensities. **e**, Mean latency difference in auditory nerve responses (left – right,  $n = 11$ ) versus speaker position. Latencies in the left auditory nerve are shorter than those of the right nerve for leftward sound source locations, but are longer for rightward locations.

of receptors with lower individual timing accuracy—a phenomenon termed temporal hyperacuity<sup>8,9</sup>. Consistently, the fly auditory system seems to be specialized for timing information, with frequency discrimination accomplished primarily through peripheral filtering mechanisms<sup>4</sup>. If receptor neurons are pooled solely to improve timing information, such a system only requires the accurate preservation of the relative timing of inputs from the two ears at subsequent stages of processing<sup>16</sup>. Through this economy of coding, the relatively simple nervous system of the fly can show hyperacute sensitivity.

Our results provide a neuroethological analysis of a remarkable behavioural ability that links both applied and theoretical studies of sensory systems. These results have significant implications for efforts to develop artificial acoustic devices based on the design principles of *O. ochracea* ears<sup>6</sup>. The acoustic orientation behaviour of *O. ochracea* demonstrates that such devices have the potential for very high sensitivity and accuracy. However, the flies' performance is not based solely on biomechanics (that is, the inherent directional properties of its eardrums). The associated neural apparatus is also highly specialized for this singular task. Directional information, initially encoded by *O. ochracea*'s tympanal mechanics, is represented by a population latency code at the receptor level, and then translated into behavioural output with high efficiency and accuracy. *Ormia ochracea* is an ideal system for investigating the statistics of neural information processing<sup>17</sup>. Its behavioural output is tightly linked to sensory input by the primary receptors, and receptor

coding is accessible to detailed study at both single-cell and population levels. Although subsequent processing in the fly's central nervous system and motor system is necessary to generate the act of steered locomotion, our results show that for the task of sound localization, the afferent code is simple and well defined. The fly's specializations in auditory processing are as remarkable as its unusual tympanal apparatus. □

## Methods

### Behavioural measurements

We measured auditory directionality using phonotactic walking responses (*O. ochracea* make their final approach to a singing cricket by walking). Flies were fixed to a wire with low-melting-point wax, and this wire was attached to a micromanipulator. Under red light, mounted flies were then placed in a normal walking position on a spherical treadmill. The treadmill consisted of an optically actuated computer-pointing device (Logitech Marble Mouse) that was modified to hold a table-tennis ball floating on an air stream. A random dot pattern on the table-tennis ball activated the optical sensor when the ball was rotated by the walking movements of the tethered fly (for a video clip of the fly's response see [www.scar.utoronto.ca/~amazon/Movie/flyball2.mpg](http://www.scar.utoronto.ca/~amazon/Movie/flyball2.mpg)). The fly's virtual trajectory was recorded by computer, using custom software that captured coordinates at 0.5 ms intervals. The treadmill was located at the centre of rotation of a speaker (Sony MDR-ED228LP, ~1 cm diameter in a 3 cm baffle) that was attached to a movable arm at a distance of 12 cm from the position of the fly. The speaker could be rotated in  $0.5^\circ$  measurable increments through  $40^\circ$  on either side of the midline axis of the fly. We synthesized acoustic stimuli using Tucker-Davis Technologies (TDT) System II hardware and custom software. Stimuli consisted of a train of 5-kHz, 10-ms tone pulses delivered at a rate of 50 pulses  $s^{-1}$  with 10 pulses per train for a total stimulus duration of 200 ms. Stimuli were amplified (Harman Kardon PM655), passed through a computer-controlled attenuator (TDT PA-4) and then sent to the speaker. Stimulus timing and amplitude were

controlled by computer.

For comparison of response latencies, freely walking flies were recorded using a Redlake Motionscope high-speed video recorder at 1,000 frames  $s^{-1}$ . Stimulus generation was as described above.

## Localization and lateralization tasks

Stimulus duration was constant across all presentations, and the response of the fly did not modify the stimulus in any way (that is, the task was open loop). Unlike closed-loop stimulus conditions, in which the directional stimulus would decrease as the fly approached the angle of the speaker, here the angle of the speaker relative to the fly was fixed, and the fly received a constant turn signal throughout the stimulus. Therefore the fly's movements that generate the measured turn angles in these experiments could not bring them to face the direction of the speaker (turn angles should overshoot the actual speaker position). If the flies were simply lateralizing the stimulus source they would receive a constant turn signal (right or left) for the duration of the stimulus, and the angle of the flies' paths relative to the midline axis should be similar for all speaker positions. If the flies were truly localizing the sound source then paths for different speaker positions should be consistently different from one another and turn angle should increase with speaker angle.

## Neurophysiology

Intracellular recordings were obtained from single auditory receptors using glass microelectrodes (borosilicate, thin-walled 1.0 mm o.d., 70–120 M $\Omega$ ). Receptors were recorded in the pterothoracic ganglion near the point of entry of the frontal nerve, which carries the auditory axons. After recordings, cells were injected with Lucifer Yellow and visualized under fluorescence microscopy to confirm that recordings were from primary receptors. Amplified (AM Systems Model 1600) neural responses were recorded by computer (100-kHz sampling rate) using an analogue-to-digital interface (TDT AD1). Stimuli were tone pulses of varying frequency and duration generated using the same system as in the behavioural experiments, but delivered through a different speaker (Realistic Horn Tweeter).

We obtained auditory nerve recordings using tungsten wire electrodes (AM Systems, 0.25 mm). Amplified (AM Systems Model 1800), summed action potentials from both nerves were recorded by computer (100-kHz sampling rate). We carried out recording and stimulation as described above. For these recordings, the preparation was mounted in the same apparatus that was used for the treadmill experiments, and responses from the two auditory nerves were recorded simultaneously for stimuli delivered from different angles of incidence. As the sampling rate of our acquisition system was limited to 100 kHz, the smallest angle of incidence measured in these experiments was 10°.

Received 4 September; accepted 20 December 2000.

- Michelsen, A. in *The Evolutionary Biology of Hearing* (eds Webster, D. B., Fay, R. R. & Popper, A. N.) 61–78 (Springer, New York, 1992).
- Robert, D., Amoroso, J. & Hoy, R. R. The evolutionary convergence of hearing in a parasitoid fly and its cricket host. *Science* **258**, 1135–1137 (1992).
- Robert, D., Miles, R. N. & Hoy, R. R. Tympanal mechanics in the parasitoid fly *Ormia ochracea*: intertympanal coupling during mechanical vibration. *J. Comp. Physiol.* **183**, 443–452 (1998).
- Miles, R. N., Robert, D. & Hoy, R. R. Mechanically coupled ears for directional hearing in the parasitoid fly *Ormia ochracea*. *J. Acoust. Soc. Am.* **98**, 3059–3070 (1995).
- Robert, D., Read, M. P. & Hoy, R. R. Tympanal hearing in tachinid flies (Diptera, Tachinidae, Ormiini): The comparative morphology of an innovation. *Cell Tissue Res.* **284**, 435–448 (1996).
- Gibbons, C. & Miles, R. N. in *Proceedings of the IMECE 5–20* (ASME, Orlando, Florida, 2000).
- Fay, R. R. *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, Illinois, 1988).
- Rose, G. & Heiligenberg, W. Temporal hyperacuity in electric fish. *Nature* **318**, 178–180 (1985).
- Kawasaki, M., Rose, G. & Heiligenberg, W. Temporal hyperacuity in single neurons of electric fish. *Nature* **336**, 173–176 (1988).
- Walker, T. J. Phonotaxis in female *Ormia ochracea* (Diptera: Tachinidae), a parasitoid of field crickets. *J. Insect Behav.* **6**, 389–410 (1993).
- Cade, W. H., Ciceran, M. & Murray, A. M. Temporal patterns of parasitoid fly (*Ormia ochracea*) attraction to field cricket song (*Gryllus integer*). *Can. J. Zool.* **74**, 393–395 (1996).
- Ramsauer, N. *Phonotactic Behavior of Ormia ochracea* (Diptera: Tachinidae, Ormini) in Controlled Acoustic Conditions. Thesis, Univ. Zurich (1999).
- Robert, D. & Willi, U. The histological architecture of the auditory organs in the parasitoid fly *Ormia ochracea*. *Cell Tiss. Res.* **301**, 447–457 (2000).
- Bair, W. & Koch, C. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput.* **8**, 1185–1202 (1996).
- Carr, C. E., Heiligenberg, W. & Rose, G. J. A time-comparison circuit in the electric fish midbrain. I. Behavior and physiology. *J. Neurosci.* **6**, 107–119 (1986).
- Maršálek, P., Koch, C. & Maunsell, J. On the relationship between synaptic input and spike output jitter in individual neurons. *Proc. Natl Acad. Sci. USA* **94**, 735–740 (1997).
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. R. & Bialek, W. *Spikes* (MIT, Cambridge, Massachusetts, 1997).
- Batschelet, E. *Circular Statistics in Biology* (Academic, New York, 1981).

## Acknowledgements

We thank T. Adelman, D. Bodnar, B. Land, R. Wyttenbach, M. Andrade and all participants of NEJC. This work was supported by an NIH grant to R.R.H.

Correspondence and requests for materials should be addressed to A.C.M. (e-mail: amason@scar.utoronto.ca).

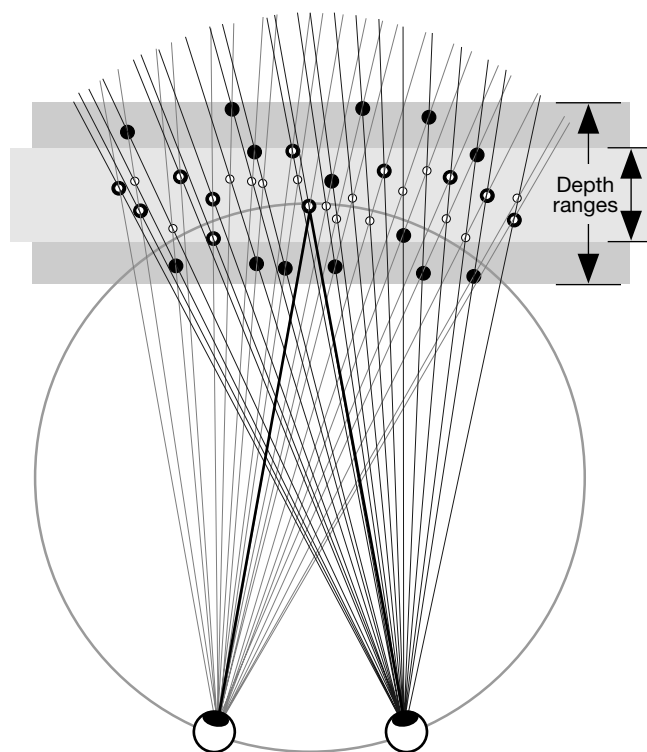
# Motion direction, speed and orientation in binocular matching

Raymond van Ee\*† & Barton L. Anderson\*

\* MIT Department of Brain and Cognitive Sciences, Cambridge, Massachusetts 02139, USA

† Utrecht University, Helmholtz Institute, Princeton Plein 5, 3584CC, Utrecht, The Netherlands

The spatial differences between the images seen by the two eyes, called binocular disparities, can be used to recover the volumetric (three-dimensional) aspects of a scene. The computation of disparity depends upon the correct identification of corresponding features in the two images. Understanding what image features are used by the brain to solve this matching problem is one of the main issues in stereoscopic vision<sup>1</sup>. Many cortical neurons in visual areas V1 (ref. 2), MT (refs 3, 4) and MST (refs 5, 6) that are tuned to binocular disparity are also tuned to orientation, motion direction and speed. Although psychophysical work has shown that motion direction<sup>7</sup> can facilitate binocular matching, the psychophysical literature on the role of orientation is mixed<sup>8,9</sup>, and it has been argued that speed differences are ineffective in aiding correspondence<sup>7</sup>. Here we use a different psychophysical paradigm to show that the visual system uses similarities in orientation, motion direction and speed to achieve binocular correspondence. These results indicate that cells that multiplex orientation, motion direction, speed and



**Figure 1** Top view of the geometry of the binocular matching problem. The intersections of the visual lines indicate the locations of possible matches. The drawing is schematic; contours in three dimensions introduce another dimension of ambiguity (namely, determining how correspondence between portions of the contours in the two eyes should be defined). One configuration represents a large depth range (black dots), the other a small depth range (white dots). The two configurations create identical sets of visual lines.



# binocular disparity may help to solve the binocular matching problem.

When two arbitrary, uncorrelated images are projected to the two eyes, very little volumetric depth is experienced. We reasoned that, if a stimulus feature is involved in the matching of stimuli from the two eyes for depth perception, the range of perceived depth in complex stereoscopic displays will be greater when this feature is present than when it is absent (Fig. 1). To investigate the role of orientation, motion direction and speed in binocular matching, we generated a stimulus that allowed us to manipulate these stimulus dimensions independently. The stimulus consisted of an array of bars (Fig. 2a) uniformly distributed inside a three-dimensional volume. The bars were arranged in a series of frontoparallel planes, and their orientation, motion direction and speed were varied. Subjects judged the perceived depth-to-width ratio (Fig. 2b) of the volume of bars by adjusting the aspect ratio of an outline rectangle (Fig. 2c). In simple configurations containing only two oblique bars, very few false matches are possible, and the relative depth of the bars is readily recovered in both static and motion conditions. However, when many static bars with similar orientations are visible, more false matches are possible, and observers report a substantial reduction in their overall impression of depth. This reduction in depth presumably reflects a bias to match features that lie near the horopter, in other words those that have zero disparity. In principle, matching ambiguity could be considerably reduced if the visual system could use the motion direction, speed or orientation of the bars, or a combination of these features.

To investigate this possibility, we compared the perceived depth range of orientated bars that were presented either statically or with a gradient of motion (motion parallax). In the motion parallax condition, all of the bars moved simultaneously in the same direction and the speed of a bar decreased as the simulated distance from the observer increased. Perceived depth-to-width ratios for one subject are shown in Fig. 3a. These data show that: the perceived depth of the volume decreases as the number of bars increases; the fall-off in perceived depth is less when the range of orientations increases; and the perceived depth of the volume is greater with moving than with static bars.

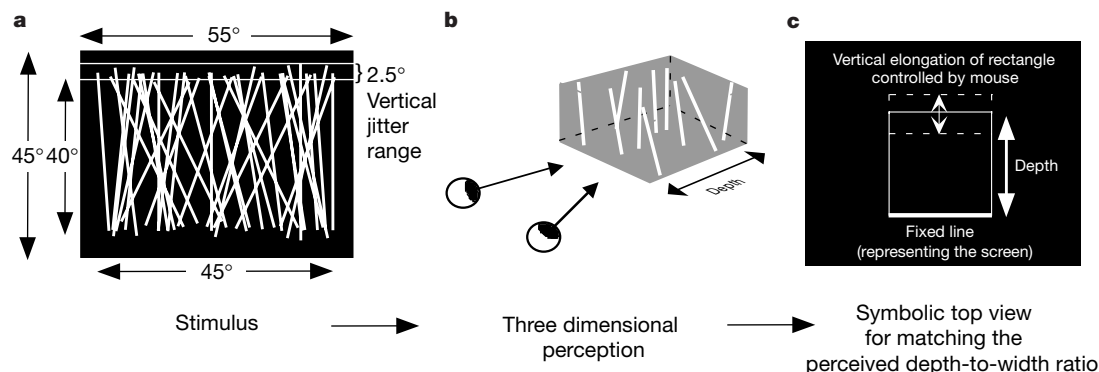
The most interesting measure is the difference between the motion and the static conditions. Figure 3a (bottom row) and 3b shows this difference for one subject and the mean of the differences for three subjects, respectively. Note that each of the panels contains a local peak that shifts gradually to the right when the range of bar orientations increases. The perceived depth-to-width ratio as a function of orientation for an intermediate bar density (140 bars) is shown in Fig. 3c for the static display. The slope of the linear fit is significantly different from zero ( $P < 0.003$ ,  $r^2 = 0.91$ ), indicating that orientation, like motion, can disambiguate matching.

The results of experiment 1 revealed that both contour orienta-

tion and motion parallax enhance stereoscopically perceived depth. However, fixation was unrestricted, so it is not clear whether the benefit of motion arose from variations in speed, direction or both. We therefore tested four new motion conditions consisting of the possible permutations of the pairs (motion direction, speed) and (varying, static across bars). Fixation was restricted to dissociate the effects of direction and speed. To maximize our chances of observing significant differences, we used a bar density and orientation range that showed optimal depth facilitation of motion in experiment 1. Icons representing the bar motions are presented along the abscissa of the histogram in Fig. 4a for the six motion conditions. In the motion parallax condition (the leftmost panel), speed varied but motion direction was constant. The second condition resembled the motion parallax condition, but the depth of individual bars as specified by disparity was uncorrelated with the depth specified by motion parallax. Note that monocularly, the second condition is indistinguishable from the motion parallax condition, but binocularly, it generates a cue conflict. The conflict arises because bars with the same speed appear in a single depth plane when viewed monocularly, but appear in different depth planes when viewed binocularly (assuming that correct matches have been identified). In the third condition, bar motion was independent of the motion direction and the speed of neighbouring bars. The only constraint was that the same collection of speeds was used in this motion condition as in the first two conditions. In the fourth condition, the speed of individual bars was fixed ( $4.4^\circ \text{ s}^{-1}$ ) but half of the bars moved to the right and the other half to the left. In the fifth motion condition, all the bars moved in the same direction with identical magnitudes. In this condition there was no differential motion. The rightmost icon represents the static condition.

The mean perceived depth-to-width ratios of experiment 2 are shown in Fig. 4a. Perceived depth in the static condition was much smaller than in the parallax condition, replicating the motion results of experiment 1. The critical data of this graph are depicted in the leftmost four panels, which show that the binocular conditions containing a range of speeds, motion directions or both are all effective in eliciting a percept of volumetric depth beyond that generated by either monocular motion or binocular disparity alone. These results show that speed, not just motion direction, can be used to disambiguate stereoscopic matches.

Our results contrast with a report that failed to find a benefit of speed in binocular matching<sup>7</sup>. To determine the cause of this difference, we attempted to replicate the results of that study using our method. We therefore constructed a random-dot display that contained only two motions and a small disparity difference between the two layers (4 arcmin). In this parameter regime, no significant differences in perceived depth were observed in the monocular and binocular displays containing only speed differences, but a large difference was observed for the display containing



**Figure 2** Experimental procedure. The stimulus consisted of a collection of frontoparallel bars (a) with disparities specifying a volume that evoked a percept of a three-dimensional

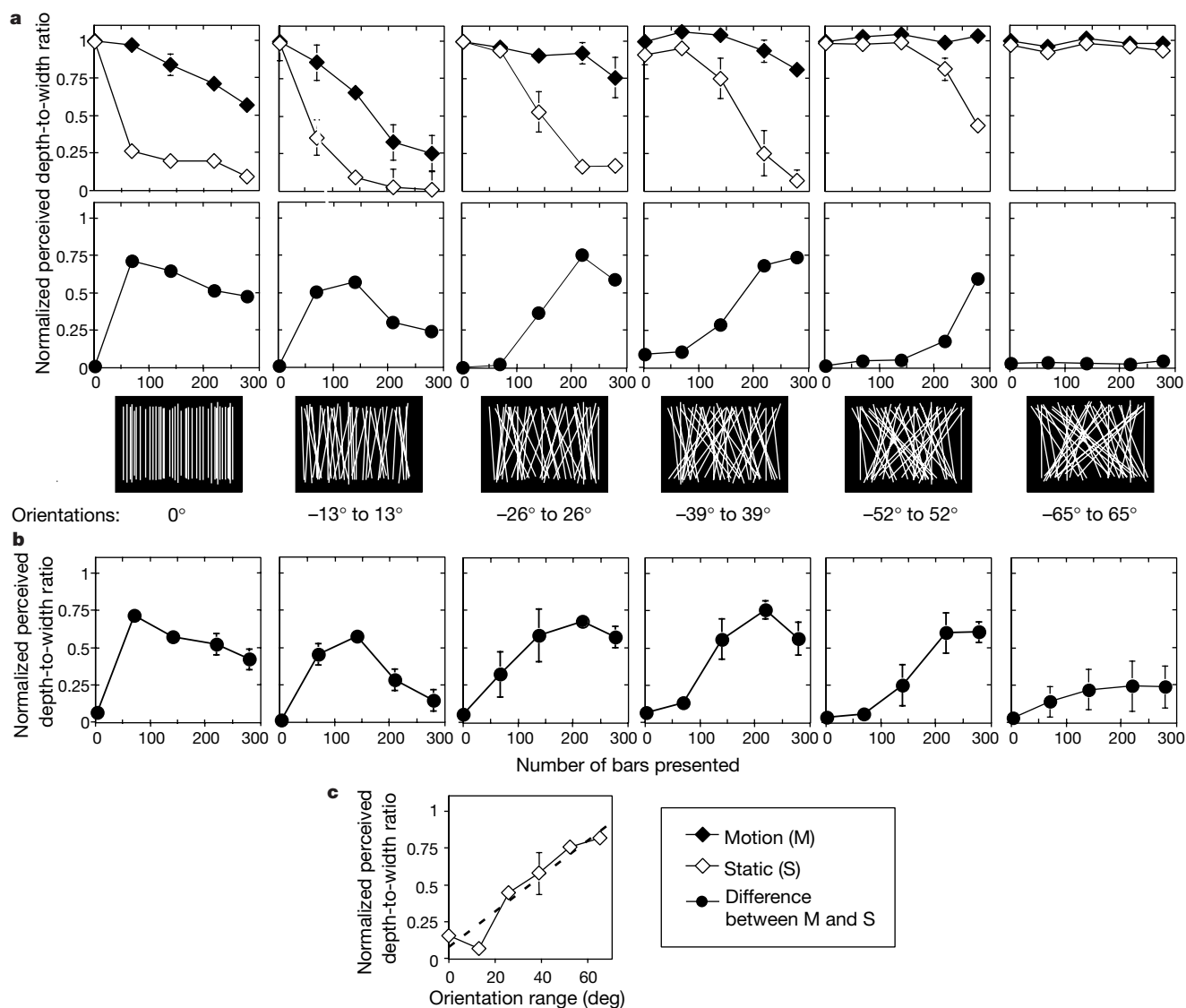
volume of bars (b). Observers matched the perceived depth-to-width ratio evoked by the stimulus with the aspect ratio of the outline rectangle (c).

different motion directions. However, the monocular percept of depth was much larger in the display containing only speed differences than in the display containing only direction differences. This suggests that the failure to obtain a benefit from speed in this parameter regime may be due to a Weber-law-like limitation on threshold sensitivity, wherein the ability to detect a just-noticeable difference scales with the standard against which it is compared. Indeed, when a suprathreshold disparity between the two layers of dots is introduced (30 arcmin), a clear benefit of speed is observed (Fig. 4c).

It is generally appreciated that disparity and motion parallax are powerful cues for the recovery of the volumetric aspects of a scene (for example, refs 10–13). It has been suggested that motion parallax and stereopsis generate independent depth estimates that are subsequently averaged to elicit an overall impression of depth<sup>14</sup>. Such a cue combination cannot account for our results. The perceived depth reported when motion and disparity were simultaneously present was larger than any weighted average of the depth elicited by these two cues when they were presented alone, even

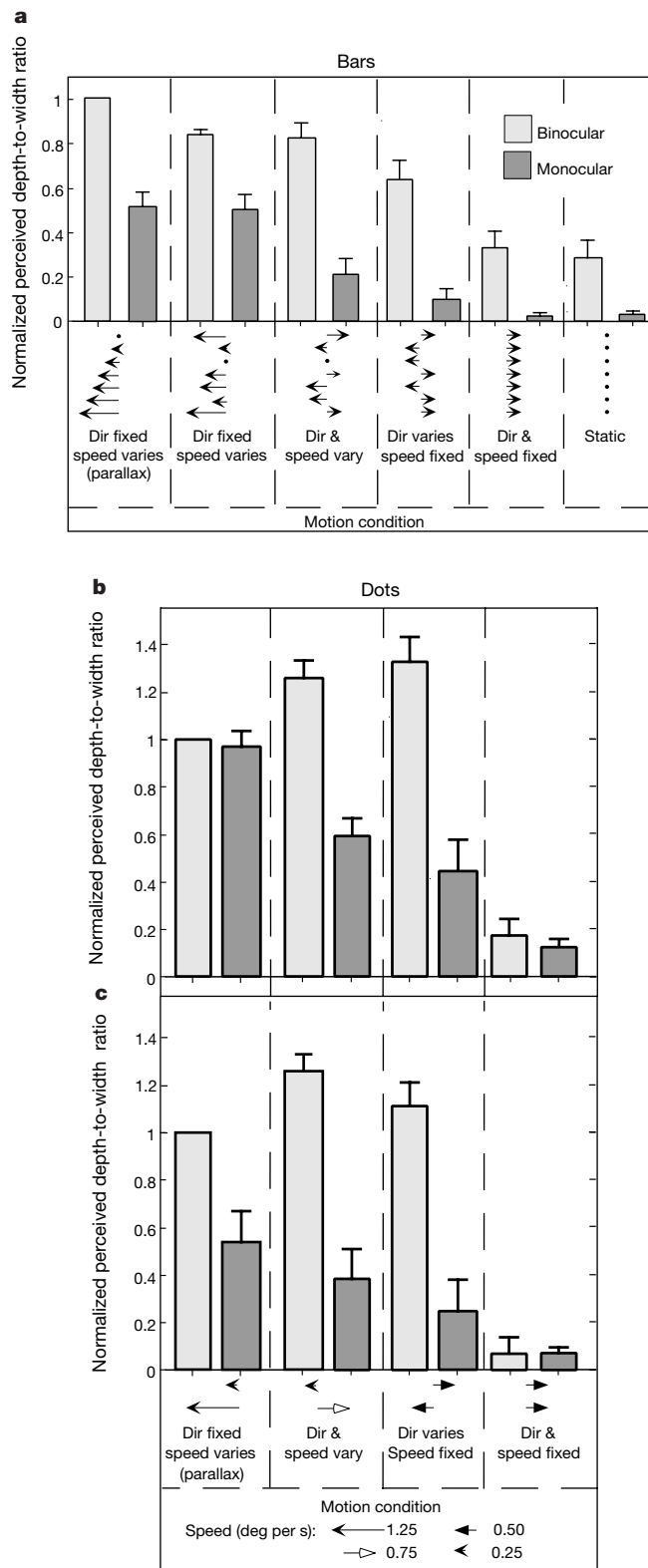
when they were in conflict (second panel of Fig. 4a). This suggests that the primary role of motion, like orientation, is to provide information that can be used to establish binocular correspondence. This interpretation is also consistent with the phenomenological reports of observers. When little depth was experienced, the images took on the appearance of an uncorrelated stereogram, generating only a reduced impression of depth generated by the few spurious matches that formed near the horopter.

Since the work of Julesz<sup>15</sup>, one of the most fundamental and unresolved problems in stereoscopic research has concerned the primitives used to establish correspondence. Resolving this problem is critical for both computational theory and neurophysiological investigations into stereopsis. In contrast to earlier work that suggested that the matching primitives are unorientated<sup>8</sup>, our results support more recent findings that the mechanisms mediating human stereopsis are orientationally tuned<sup>9</sup>. In addition, our results show that both the direction and speed of moving targets can be used to resolve matching ambiguities. Many cells in V1 (ref. 2), MT (refs 3, 4) and MST (refs 5, 6) that are tuned to binocular



**Figure 3** Results of experiment 1 showing the role of bar orientation and bar number. The icons in the centre depict the orientation range. **a**, Normalized perceived depth ratio as a function of the number of bars in the static and motion parallax conditions for observer J.E. (data from other observers are similar). The perceived depth ratio in the motion condition when only two bars were present was normalized to unity and all other depth

ratios were scaled by the same factor. Filled circles, differences in perceived depth ratio between the two conditions. **b**, Mean difference in perceived depth ratio between the static and motion parallax conditions across three subjects. Error bars,  $\pm 1$  s.d. **c**, Mean perceived depth-to-width ratio as a function of orientation for 140 static bars. Dashed line, a linear fit to the data ( $r^2 = 0.91$ ).



**Figure 4** Results of experiments 2 and 3 showing the roles of motion direction and speed. The stimuli consisted of bars and dots, respectively. The motion conditions are depicted along the abscissa of the histograms. The perceived depth ratios in the motion parallax and stereo condition were normalized to unity and all other conditions were scaled accordingly. Dir, direction. **a**, Both speed and motion differences enhance perceived depth. **b**, Replication of ref. 7. For small disparity differences (4 arcmin), there is no clear benefit of speed. **c**, Replication of experiment 2. When the disparity difference between the two depth planes is sufficiently large (30 arcmin), a substantial benefit of speed is observed. Error bars, s.e. across three observers.

disparity are also tuned to orientation, motion direction and speed. Our results indicate that one of the functional roles of cells that multiplex motion information and orientation may be to facilitate binocular matching, shedding new light on the neural substrates that subserve stereopsis in primates. □

## Methods

The stimuli were anaglyph stereograms containing bright frontoparallel bars (10.8 arcmin widths) on a black background that were rear-projected onto a large flat screen. The displays were viewed from a distance of 2 m. The bars were presented behind the screen to avoid conflicting occlusion cues along the screen boundaries (note that any bar that intersected the left or right side of the display would monocularly appear to be occluded by the border of the screen). The far end of the volume had a relative disparity of  $1.6^\circ$ . The locations of bar end points were randomly chosen from a vertical range of  $2.5^\circ$  (independently of their orientation; Fig. 2a). The bars were randomly positioned within the volume such that the disparity steps between the bars was  $1.6^\circ$  divided by the number of bars. Overlap in the bars was possible due to the random positioning. The minimal and maximal values of disparity ( $0^\circ$  and  $1.6^\circ$ ) were present in all displays.

Subjects judged the perceived depth-to-width ratio of the volume by matching this to the aspect ratio of a rectangular display (Fig. 2c). Although the disparity range used was consistent with a volume that receded to 14 m, observers reported that the perceived depth was almost always smaller than the perceived width of the volume (and never exceeded the values available in the rectangular display). In each block of trials, all of the stimulus conditions appeared once in random order for 3.0 s. In each experiment there were ten trial blocks. In the motion conditions, the bars underwent a square-wave oscillation (left-right in a frontoparallel plane), reversing direction three times (two cycles). The bar speeds spanned the range of  $1.2^\circ \text{ s}^{-1}$  to  $7.6^\circ \text{ s}^{-1}$  (0.5 cycle was spanned by 20 frames). Observers R.E., R.F., J.E. and E.B. participated in experiment 1, and R.E., J.E. and E.B. participated in experiments 2 and 3.

## Experiment 1

The stimuli subtended  $45^\circ$  horizontally by  $40^\circ$  vertically. The number of bars (2, 70, 140, 210 or 280) and their orientations varied across trials. Bar orientations were randomly chosen from a range of  $\pm 0, 13, 26, 39, 52$  or  $65^\circ$ , centred about the vertical. The bars were presented either statically or with a gradient of motion, such that the speed decreased systematically with the depth specified by the bar's disparity (motion parallax). Eye movements were unrestricted. Each observer participated in 600 trials (5 numbers of bars, 6 bar orientation ranges, 2 motion conditions, 10 trial repetitions).

## Experiment 2

We investigated static and motion parallax displays as well as four other motion conditions. These conditions consisted of the possible permutations of the pairs (direction, speed) and (varying, static across bars). Motion was again in the frontoparallel plane. Details about the motion conditions are provided in the main text. Viewing was either monocular or binocular. Observers were instructed to fixate a symbol to dissociate the effects of the direction and speed components of motion. The fixation symbol was an outline square ( $1.8^\circ \times 1.8^\circ$ ) with line segment widths of 5.4 arcmin. It was always presented in the plane of the screen (zero disparity). Under restricted fixation conditions, pilot work revealed that perceived depth was more vivid when a smaller stimulus size ( $20^\circ \times 20^\circ$ ) was used. We used a bar density and orientation that showed optimal facilitation of motion on perceived depth in experiment 1: there were 62 bars (the same density as resulted from 140 bars in experiment 1) and the orientation range was  $\pm 26^\circ$ . Thus, experiment 2 consisted of 120 trials (6 motion conditions, 2 viewing conditions, 10 trial repetitions).

## Experiment 3

This experiment was identical to experiment 2, with the following modifications: 850 dots ( $5.4 \times 5.4 \text{ arcmin}^2$ ) were used instead of bars; only two layers were simulated; and the display size was reduced to  $8^\circ \times 5^\circ$ . The parametric changes were introduced to simulate more closely the conditions studied in ref. 7. The nearest layer had zero disparity. The more distant layer had a disparity of 4 arcmin in the first session and 30 arcmin in the second. The motion direction and speed were uniform within each layer. The permutations of the pairs (motion direction, speed) and (varying, static) were tested. Dot speeds ( $0.25, 0.50, 0.75$  or  $1.25^\circ \text{ s}^{-1}$ ) were chosen such that the differential speed between the two layers was  $1^\circ \text{ s}^{-1}$ .

Received 16 August 2000; accepted 14 January 2001.

- Howard, I. P. & Rogers, B. J. *Binocular Vision and Stereopsis* (Oxford Univ. Press, New York, 1995).
- Poggio, G. F. & Talbot, W. H. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J. Physiol.* **315**, 469–492 (1981).
- Bradley, D. C., Qian, N. & Andersen, R. A. Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* **373**, 609–611 (1995).
- Maunsell, J. H. & Van Essen, D. C. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* **49**, 1148–1167 (1983).
- Roy, J. P. & Wurtz, R. H. The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature* **348**, 160–162 (1990).
- Roy, J. P., Komatsu, H. & Wurtz, R. H. Disparity sensitivity of neurons in monkey extrastriate area MST. *J. Neurosci.* **12**, 2478–2492 (1992).



7. Bradshaw, M. F. & Cumming, B. G. The direction of retinal motion facilitates binocular stereopsis. *Proc. R. Soc. Lond. B* **264**, 1421–1427 (1997).
8. Mayhew, J. E. & Frisby, J. P. Stereopsis masking in humans is not orientationally tuned. *Perception* **7**, 431–436 (1978).
9. Mansfield, J. S. & Parker, A. J. An orientation-tuned component in the contrast masking of stereopsis. *Vision Res.* **33**, 1535–1544 (1993).
10. Rogers, B. J. & Graham, M. E. Similarities between motion parallax and stereopsis in human depth perception. *Vision Res.* **22**, 261–270 (1982).
11. Bradshaw, M. F. & Rogers, B. J. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Res.* **36**, 3457–3468 (1996).
12. Tittle, J. S. & Braunstein, M. L. Recovery of 3-D shape from binocular disparity and structure from motion. *Percept. Psychophys.* **54**, 157–169 (1993).
13. Richards, W. Structure from stereo and motion. *J. Opt. Soc. Am. A* **2**, 343–349 (1985).
14. Landy, M. S., Maloney, L. T., Johnston, E. B. & Young, M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* **35**, 389–412 (1995).
15. Julesz, B. *Foundations of Cyclopean Perception* (Univ. Chicago Press, Chicago, 1971).

## Acknowledgements

R.v.E. was supported by a NIH grant awarded to B.A. and by the Royal Netherlands Academy of Arts and Sciences, and B.A. was supported in part by NIH.

Correspondence and requests for materials should be addressed to B.L.A. (e-mail: bart@mit.edu).

## The homeobox gene *lim-6* is required for distinct chemosensory representations in *C. elegans*

Jonathan T. Pierce-Shimomura, Serge Faumont, Michelle R. Gaston, Bret J. Pearson & Shawn R. Lockery

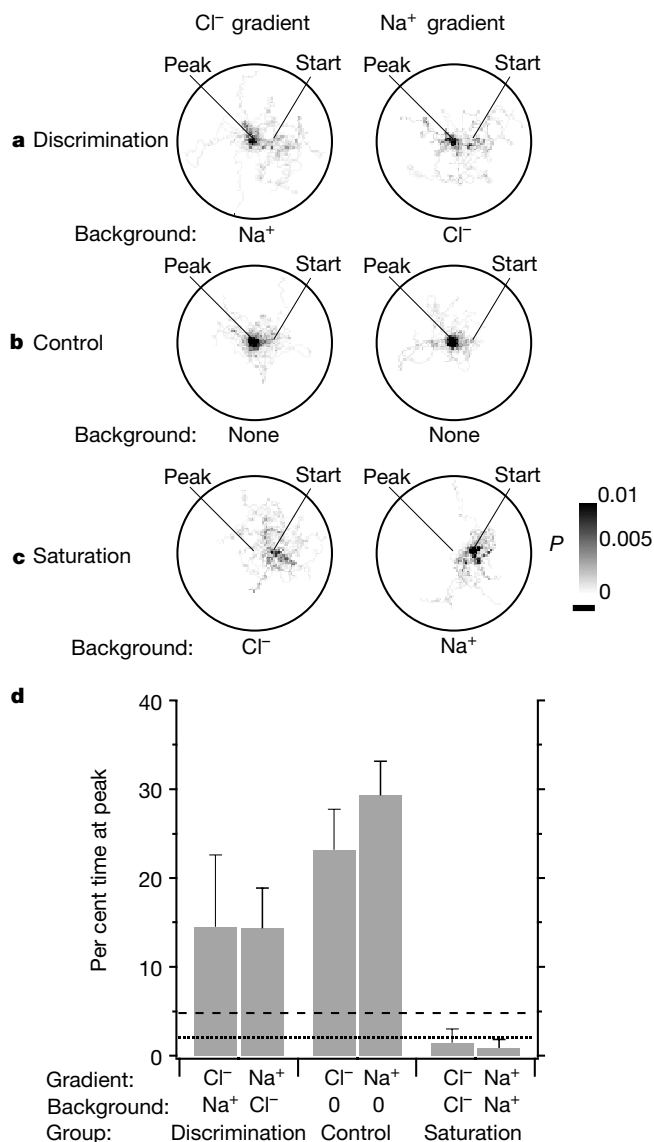
Institute of Neuroscience, 1254 University of Oregon, Eugene, Oregon 97403, USA

The ability to discriminate between different chemical stimuli is crucial for food detection, spatial orientation and other adaptive behaviours in animals. In the nematode *Caenorhabditis elegans*, spatial orientation in gradients of soluble chemoattractants (chemotaxis) is controlled mainly by a single pair of chemosensory neurons<sup>1</sup>. These two neurons, ASEL and ASER, are left–right homologues in terms of the disposition of their somata and processes, morphology of specialized sensory endings, synaptic partners and expression profile of many genes<sup>2,3</sup>. However, recent gene-expression studies have revealed unexpected asymmetries between ASEL and ASER. ASEL expresses the putative receptor guanylyl cyclase genes *gcy-6* and *gcy-7*, whereas ASER expresses *gcy-5* (ref. 4). In addition, only ASEL expresses the homeobox gene *lim-6*, an orthologue of the human *LMX1* subfamily of homeobox genes<sup>5</sup>. Here we show, using laser ablation of neurons and whole-cell patch-clamp electrophysiology, that the asymmetries between ASEL and ASER extend to the functional level. ASEL is primarily sensitive to sodium, whereas ASER is primarily sensitive to chloride and potassium. Furthermore, we find that *lim-6* is required for this functional asymmetry and for the ability to distinguish sodium from chloride. Thus, a homeobox gene increases the representational capacity of the nervous system by establishing asymmetric functions in a bilaterally symmetrical neuron pair.

To determine whether *C. elegans* can distinguish between the attractants sodium and chloride in our experimental system<sup>6</sup>, we performed a discrimination test in which we examined the chemotaxis performance of worms in a gradient of sodium (50  $\mu$ M to 10 mM) superimposed on a saturating background concentration of chloride (100 mM), and vice versa (Fig. 1a, d). In both cases, worms migrated toward the peak of the gradient as expected<sup>7</sup>,

although the overall chemotaxis performance was reduced relative to control worms tested in sodium or chloride gradients but with no background ion concentration (Fig. 1b, d).

However, the chemotaxis performance of worms tested under conditions in which the gradient and background concentration contained the same ion (Fig. 1c, d) degenerated to roughly chance level, defined as the chemotaxis performance of worms tested in the absence of a gradient (Fig. 1d, dotted line). Together, these results show that sodium chemotaxis persists against a background of



**Figure 1** Discrimination of sodium and chloride by wild-type *C. elegans*. **a–c**, Probability density plots for worms assayed in gradients of chloride (left) and sodium (right) originating at the centre of the plate (peak),  $n \geq 15$ . Grey scale indicates the probability per unit area of finding a worm at a given location in the plate during the 20-min assay. Scale bar, 1 cm. **a**, Discrimination. Worms located and remained at the peak of a chloride gradient superimposed on a high background concentration (100 mM) of sodium, or at the peak of a sodium gradient superimposed on a high background concentration (100 mM) of chloride. **b**, Control. Chemotaxis performance was normal in gradients identical to those in **a** when the background ion was omitted. **c**, Saturation. Worms failed to locate the peak of the gradient when it was superimposed on a high background concentration (100 mM) of the same ion. **d**, Average per cent time spent at the gradient peak for worms in the six conditions shown in **a–c**. Dotted line indicates mean chance performance, measured by assaying 74 worms in plates with neither gradient nor background. Dashed line indicates upper 95 per cent confidence limit for mean chance performance.

chloride sufficient to saturate the chloride response, and chloride chemotaxis persists against a background of sodium sufficient to saturate the sodium response. Thus, *C. elegans* has distinct sensory pathways for detecting sodium and chloride ions, in agreement with previous qualitative observations<sup>7</sup>.

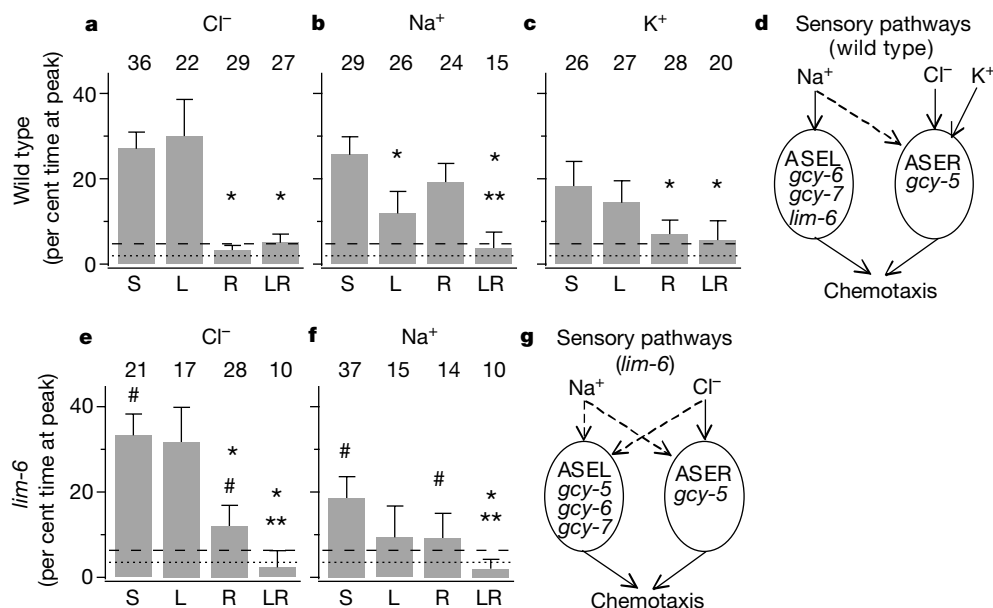
To identify the neuronal constituents of these two sensory pathways, we compared the effects on chloride and sodium chemotaxis of unilateral ablations of the ASE neurons (Fig. 2a, b). We found that chloride chemotaxis was disrupted by ablation of ASER, indicating that it is a necessary constituent of the chloride sensory pathway. In contrast, chloride chemotaxis was unaffected by ablation of ASEL, suggesting that ASEL is not a necessary constituent of the chloride pathway. Furthermore, we found that sodium chemotaxis was disrupted by ablation of ASEL, indicating that it is a necessary constituent of the sodium sensory pathway. Although ablation of ASER had little or no effect on sodium chemotaxis, the effect of bilateral ablation of the ASE neurons was greater than the effect of unilateral ablation of ASEL (Bonferroni post hoc comparison,  $P < 0.001$ ). Thus, ASER seems to be a secondary constituent of the sodium pathway. We conclude that the asymmetries in gene expression in the ASE neurons coincide with differences in sensory function.

As the strongest unilateral ablation effects were for ASEL in the case of sodium chemotaxis and ASER in the case of chloride chemotaxis, we thought that ASEL and ASER might be specialized for detection of cations and anions, respectively. Therefore, we compared the effects of unilateral and bilateral ablation of ASE neurons on potassium chemotaxis as an example of a response to a second cation (Fig. 2c). We found that potassium chemotaxis was disrupted by ablation of ASER, but not by ablation of ASEL, indicating that ASER is a necessary constituent of the potassium pathway, but ASEL is not. Thus, ASER is not specialized for detecting anions, and the functional asymmetries between ASEL and ASER extend to a third ion.

The ablation data are summarized in Fig. 2d, which shows the respective contribution of the ASE neurons to the sodium, chloride and potassium pathways. (The three other chemosensory neuron pairs known to be necessary for normal chemotaxis to sodium and chloride ions—ADF, ASG and ASI—were not tested because their joint contribution to chemotaxis is much less than that of the ASE neurons<sup>1</sup>.) Only ASER contributes to the chloride and potassium pathways. Both ASER and ASEL contribute to the sodium pathway, but the latter makes the stronger contribution.

Our finding that ASEL, the main constituent of the sodium pathway, is not part of the potassium pathway suggests that sodium chemotaxis should persist against a high background of potassium—a prediction at odds with previous work<sup>7</sup>. In our more quantitative behavioural assay, however, sodium chemotaxis persists against a background of 100 mM potassium (per cent time at peak:  $\text{Na}^+$  gradient with  $\text{K}^+$  background,  $20.7 \pm 6.02$ ,  $n = 18$ , versus no gradient,  $1.97 \pm 1.02$ ,  $n = 74$ ;  $t = 6.02$ ,  $P < 0.0001$ ). We did, however, observe a substantial reduction in potassium chemotaxis against a background of 100 mM sodium (per cent time at peak:  $\text{K}^+$  gradient with  $\text{Na}^+$  background,  $8.37 \pm 5.35$ ,  $n = 33$ ;  $\text{K}^+$  gradient with no background,  $23.1 \pm 5.67$ ,  $n = 15$ ;  $t = 3.91$ ,  $P < 0.001$ ), consistent with previous results<sup>7</sup> and our finding that the potassium pathway and the secondary sodium pathway converge on ASER.

To determine whether the differences between ASEL and ASER in sensory function are reflected in their electrophysiology, we performed whole-cell voltage-clamp recordings from ASEL for comparison with previous recordings from ASER<sup>8</sup>. Net membrane current evoked by a family of voltage pulses between  $-154$  and  $+46$  mV (holding potential =  $-74$  mV) was qualitatively similar in ASEL and ASER (Fig. 3a). Both neurons exhibited a time-independent inward current activated by hyperpolarization below  $-90$  mV, and a fast activating outward current activated by depolarization. Outward current decayed exponentially over time in both neurons.



**Figure 2** Effect of ablations of ASEL and ASER on chemotaxis performance in chloride, sodium and potassium gradients. **a–c, e, f**, Per cent time at the gradient peak for four groups of worms: unablated sham (S), ASEL ablation (L), ASER ablation (R), and ASEL plus ASER ablation (LR). Single asterisks indicate a significant difference relative to sham ( $P < 0.01$ ); double asterisks indicate a significant difference relative to the ASEL ablation or the ASER ablation ( $P < 0.01$ ); hash indicates a significant difference relative to the same groups in wild-type worms ( $P < 0.05$ ). Numbers above the bars indicate sample sizes. Dotted lines indicate mean chance performance, measured by assaying worms in

plates with neither gradient nor background. Dashed lines indicate the upper 95 per cent confidence limit for chance performance. **d**, Summary of the sodium, chloride, and potassium sensory pathways based on the data in **a–c**. Dashed line indicates that ASER is a secondary constituent of the sodium pathway. **g**, Summary of the sodium and chloride sensory pathways in *lim-6* based on the data in **e** and **f**; the potassium pathway was not tested. Dashed lines indicate weakened or secondary sensory function. The asymmetric function of the ASE neurons is correlated with asymmetric expression of the receptor guanylyl cyclase genes *gcy-5*, *-6* and *-7* and the homeobox gene *lim-6* (refs 4, 5).

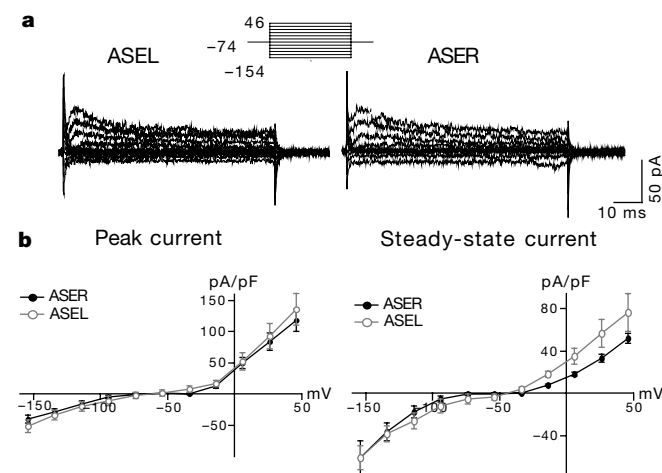
Substitution of *N*-methyl-D-glucamine for potassium in the recording pipette abolished outward current in both ASEL (data not shown) and ASER<sup>8</sup>, suggesting that it is carried by potassium ions. Rates of inactivation for outward current vary widely among *C. elegans* neurons<sup>8</sup>. We therefore compared inactivation rates of the outward current at +46 mV for ASER and ASEL. Notably, we found no statistical difference in the time constants of inactivation for ASER ( $19.7 \pm 2.4$  ms;  $n = 31$ ) and ASEL ( $20.8 \pm 3.9$  ms;  $n = 12$ ,  $t = 1.60$ ,  $P > 0.05$ ), suggesting that the transient current may be carried by the same or similar potassium channels.

Plots of peak current density versus voltage were also strikingly similar for ASEL ( $n = 14$ ) and ASER ( $n = 29$ ) (Fig. 3b, left). The plot of steady-state current density versus voltage, however, revealed that outward current density above -14 mV is increased about 1.5–2-fold in ASEL (Fig. 3b, right; analysis of variance,  $P < 0.01$ ). Thus, the difference in sensory function between ASEL and ASER is reflected in a difference in the amount of outward current generated when the neuron is depolarized, which could in turn produce a difference in the threshold for regenerative events seen in ASE neurons<sup>8</sup>. Whereas the functional consequence of this difference remains to be investigated, the overall electrophysiological similarity of the two neurons suggests that differences in the chemosensitivities of the two neurons to sodium, chloride and potassium, rather than differences in voltage-dependent currents, are responsible for the asymmetrical contributions of ASEL and ASER to chemotaxis.

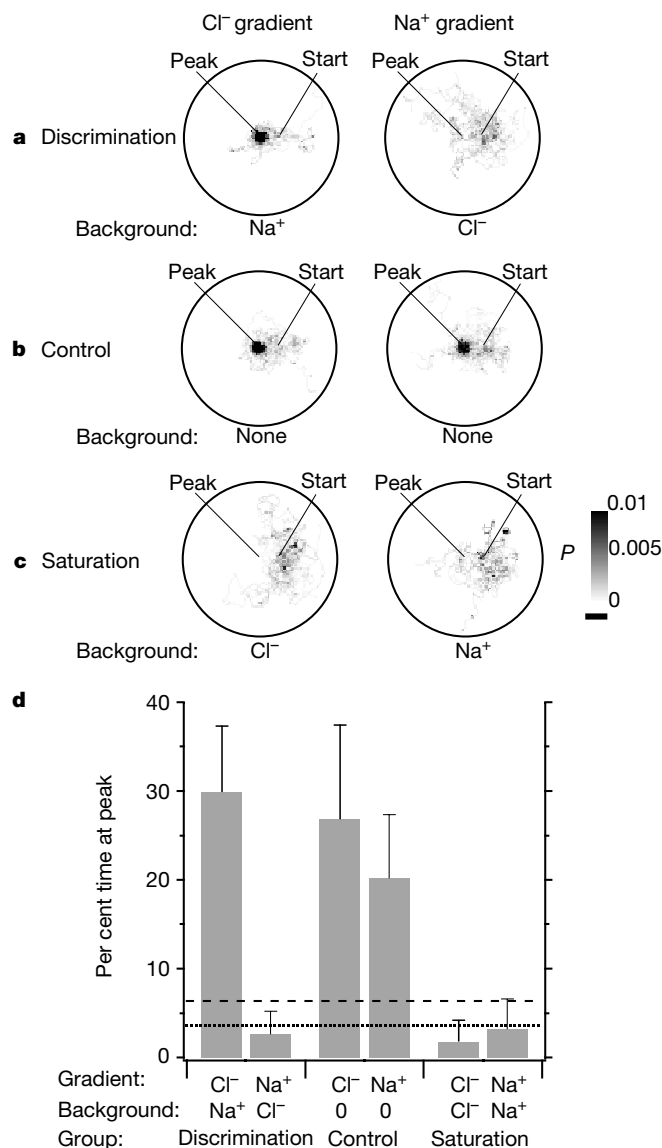
In the mutant *lim-6(nr2073)*, a putative null missing the homeo-domain and one of two LIM domains of the protein, *gcy-5*, which is normally expressed only in ASER, is also expressed in ASEL<sup>5</sup>. Thus, the *lim-6* gene is required for asymmetric *gcy-5* expression in ASEL and ASER. To determine whether the *lim-6* gene is also required for the asymmetric sensory functions of ASEL and ASER, we compared the effect on chloride chemotaxis of unilateral ablations of ASE neurons in *lim-6* mutants (Fig. 2e). We found that although chloride chemotaxis was reduced by ablation of ASER, the reduction was significantly less than in wild-type worms (Fig. 2e, R, versus Fig. 2a, R:  $t = 3.72$ ,  $P < 0.001$ ). Moreover, *lim-6* mutants in which ASER was ablated performed significantly better than chance (Fig. 2e, R, versus chance:  $t = 3.17$ ,  $P < 0.01$ ), in contrast to wild-type worms in which ASER was ablated (Fig. 2a, R, versus chance:  $t = 0.867$ ,  $P > 0.05$ ). Thus, in *lim-6* mutants, chloride

chemotaxis persists in the absence of ASER, suggesting the existence of a new constituent of the chloride pathway.

A likely candidate for this constituent was ASEL, because it now expressed *gcy-5*, a characteristic feature of the chloride-sensitive ASER neuron. Although killing ASEL had no effect on chloride chemotaxis in *lim-6* mutants, killing ASER together with ASEL reduced chloride chemotaxis more than ablating ASER alone



**Figure 3** Comparison of membrane currents in ASEL and ASER. **a**, Currents evoked by a family of 80-ms voltage pulses between -154 and +46 mV from a holding potential of -74 mV in ASEL (left) and ASER (right). **b**, Current density versus voltage plots for peak (left) and steady-state current (right) in ASER (filled circles,  $n = 29$ ) and ASEL (open circles,  $n = 14$ ). Steady-state current was calculated as the mean current over the last five milliseconds of the voltage pulse. Bars, 95 per cent confidence intervals.



**Figure 4** Discrimination of sodium and chloride in *lim-6* mutants. **a–c**, Probability density plots for *lim-6* mutants assayed in gradients of chloride (left) and sodium (right) originating at the centre of the plate (peak) ( $n \geq 15$ ). Grey scale indicates the probability per unit area of finding a worm at a given location in the plate during the 20-min assay. Scale bar, 1 cm. **a**, Discrimination. *lim-6* mutants located and remained at the peak of a chloride gradient superimposed on a high background concentration (100 mM) of sodium (left) but failed to locate the peak of a sodium gradient superimposed on a high background concentration (100 mM) of chloride (right). **b**, Control. *lim-6* mutants located and remained at the peak of gradients identical to those in **a** when the background was omitted. **c**, Saturation. *lim-6* mutants failed to locate the peak of the gradient when it was superimposed on a high background concentration (100 mM) of the same ion. **d**, Average per cent time spent at the gradient peak for worms in the six conditions shown in **a–c**. Dotted line indicates mean chance performance, measured by assaying 36 mutants in plates with neither gradient nor background. Dashed line indicates the upper 95 per cent confidence interval for chance performance.



(Bonferroni post hoc comparison,  $P < 0.001$ ), with chloride chemotaxis reduced to the chance level. These results suggest that in *lim-6* mutants ASE is responsible, at least in part, for the persistence of chloride chemotaxis in the absence of ASER. We conclude that the *lim-6* gene is required to separate fully the sensory functions of ASE and ASER.

The abnormal sensitivity of ASE to chloride in *lim-6* mutants raised the possibility that the sodium pathway was altered in these worms. Indeed, we found that sodium chemotaxis was reduced in *lim-6* mutants relative to wild-type worms (Fig. 2f, S, versus Fig. 2b, S:  $t = 2.12$ ,  $P < 0.05$ ). This reduction appears to reflect an impairment of chemotaxis ability specific to sodium because chloride chemotaxis was actually enhanced in *lim-6* mutants relative to wild-type worms (Fig. 2e, S, versus Fig. 2a, S:  $t = 2.02$ ,  $P < 0.05$ ). In contrast to unilateral ablations of the ASE neurons in wild-type animals (Fig. 2b, L and R), ablation of ASE or ASER in *lim-6* mutants produced equivalent modest, yet not significant, reductions in chemotaxis to sodium (Fig. 2f, L and R). This result suggests that in *lim-6* mutants, ASE and ASER may have identical functions in the weakened sodium pathway. Bilateral ablation of the ASE neurons resulted in a significant reduction in sodium chemotaxis (Fig. 2f, S, versus Fig. 2f, LR: Bonferroni post hoc comparison,  $P < 0.01$ ), indicating that ASE and ASER remain constituents of the sodium pathway in the *lim-6* mutant. We conclude that although the constituents of the sodium pathway were not changed in *lim-6* mutants, the *lim-6* gene is required to establish the full sensitivity to sodium of ASE. The ablation data for *lim-6* mutants are summarized in Fig. 2g, in which a strengthened chloride pathway relies on both ASE and ASER, and a weakened sodium pathway also relies on both ASE and ASER.

Previous experiments have shown that the G-protein-coupled receptor STR-2 is expressed in either the left or the right member of the chemosensory neuron pair AWC<sup>9</sup>, and that this asymmetry is necessary for behavioural discrimination in olfactory chemotaxis (P. D. Wes and C. I. Bargmann, personal communication). The fact that the *lim-6* gene is required for asymmetric expression of *gcy-5* suggests that *lim-6* may similarly be required for the behavioural discrimination between sodium and chloride ions. To test this idea, we performed a discrimination test on *lim-6* mutants under conditions identical to those in Fig. 1. We found that *lim-6* mutants could perform chemotaxis to chloride in a sodium background (Fig. 4a, d), indicating that the *lim-6* gene is not required for behavioural discrimination of chloride from sodium. In contrast, we found that *lim-6* mutants could not perform chemotaxis to sodium in a chloride background (Fig. 4a, d). Note that this effect cannot be explained by a loss of sodium sensitivity because *lim-6* mutants could perform chemotaxis to an identical sodium gradient with no background ion (Fig. 4b, d), albeit at a slightly reduced level compared with wild-type worms (Fig. 1b, d;  $t = 2.12$ ,  $P < 0.05$ ). Thus, the homeobox gene *lim-6* is required to discriminate sodium from chloride.

With just eight left–right homologous pairs of chemosensory neurons<sup>1,10–12</sup> known to direct behavioural responses to at least 100 water soluble attractants and repellents<sup>7,13</sup>, *C. elegans* faces a significant limitation in its ability to differentiate chemical stimuli. Our finding of functional asymmetries between the ASE neurons suggests that *C. elegans* has adopted the unexpected solution of increasing the effective number of different types of chemosensory neurons by breaking bilateral symmetry. Similarly, only three pairs of chemosensory neurons are known to direct responses to at least 60 volatile attractants and repellents in *C. elegans*<sup>14</sup>, and functional asymmetries have been described in one of these neuron pairs as well<sup>15</sup>. An analogous limitation may exist in the mammalian olfactory system in which hundreds of thousands of different chemical stimuli are discriminated by just 2,000 glomeruli<sup>16</sup>—the main functional units of the olfactory bulb. By analogy to *C. elegans* chemosensation, discrimination in

mammalian olfaction could be enhanced by breaking bilateral symmetry in the activation of glomeruli<sup>17</sup>. □

## Methods

### Chemotaxis assays

Single adult hermaphrodites were rinsed to remove OP50 bacteria and placed on a food-less, agar-filled plate for 10–40 min to accommodate, before being transferred to a second agar-filled plate in which a radially shaped gradient of attractant was established. The agar contained (in mM): CaCl<sub>2</sub> (1), MgSO<sub>4</sub> (1) and KPO<sub>4</sub> (25); pH 6.5. A gradient of either ammonium chloride, sodium acetate or potassium acetate was established by placing a 5-μl drop of 1 M attractant in the centre of the assay plate at 18–24 h, and again at 3–4 h before each assay<sup>7</sup>. The maximum estimated attractant concentration<sup>6</sup> was ~10 mM at the centre of the assay plate, and the minimum concentration was ~50 μM at the edge of the plate. At these concentrations, neither ammonium nor acetate ions are attractive<sup>7</sup>. We recorded movement by noting the location of the worm's centroid at 1-s intervals for 20 min using an automated tracking system with 257 pixels mm<sup>-1</sup> resolution<sup>6</sup>. Each animal started 11 mm away from the gradient peak, and was tested only once.

### Laser ablation of neurons

Neurons were ablated in L1 and early L2 larvae as described<sup>18</sup>. Neurons were identified for ablation by expression of green fluorescent protein (GFP) restricted to ASER by the *gcy-5* promoter, and restricted to ASE by the *gcy-6* promoter<sup>4</sup>. The presence of GFP did not affect chemotaxis, because performance was identical for animals with and without GFP expression in ASE neurons (per cent time at peak: animals with GFP,  $27.1 \pm 3.67$ ,  $n = 38$ , versus animals without GFP,  $23.15 \pm 4.55$ ,  $n = 17$ ;  $t = 1.31$ ,  $P > 0.05$ ). Ablated animals were tested for chemotaxis 2–4 d after ablation; individuals in which GFP was still visible in the target neuron after ablation were excluded from the behavioural analysis. In sham ablations, the laser was fired near the worm without killing any cells. We took two additional measures to control for nonspecific effects. First, worms in which ASE neurons were ablated were tested for chemotaxis to diacetyl, a compound sensed by a different chemosensory neuron pair, AWA<sup>14,19</sup>. Diacetyl gradients were made by placing a 1-μl drop of diacetyl (aqueous dilution 1:100) on the assay plate ~10 min before the start of each experiment. In agreement with previous results<sup>14</sup>, animals with unilateral and bilateral ablations of ASE neurons exhibited normal diacetyl chemotaxis (per cent time at peak: sham,  $21.7 \pm 6.16$ ,  $n = 15$ ; ASE ablated,  $22.5 \pm 10.0$ ,  $n = 9$ ; ASER ablated,  $20.7 \pm 7.44$ ,  $n = 12$ ; both ASE and ASER ablated,  $24.6 \pm 5.23$ ,  $n = 10$ ;  $F = 0.161$ ,  $P > 0.05$ ). Second, we found no difference in average instantaneous speed (used as a measure of vitality) between worms in which ASE neurons were ablated and shams (speed: ablated,  $136 \pm 5.13 \mu\text{m s}^{-1}$ ,  $n = 308$ , versus sham,  $145 \pm 7.89 \mu\text{m s}^{-1}$ ,  $n = 149$ ;  $t = 1.65$ ,  $P > 0.05$ ).

### Electrophysiology

Animals were prepared and dissected as described<sup>8,20</sup>. We carried out whole-cell patch-clamp recordings using electrodes made from borosilicate glass tubing (Sutter BF120-69-10, Novato, CA) filled with (in mM): potassium gluconate (125), KCl (18), 4 NaCl (4), MgCl<sub>2</sub> (1), CaCl<sub>2</sub> (0.6), HEPES (10), EGTA (10); pH 7.2 with KOH. In some experiments N-methyl-D-glucamine was substituted for K<sup>+</sup> in the recording pipette. External saline contained (in mM): NaCl (145), KCl (5), CaCl<sub>2</sub> (1), MgCl<sub>2</sub> (5), HEPES (10), D-glucose (20); pH 7.2 with NaOH. Membrane current was amplified with a modified Axopatch 200A (Axon Instruments, Foster City, CA), filtered at 10 kHz and digitized at 25 kHz. Voltages were corrected for liquid junction potentials. Neurons were identified by GFP expression as described above.

### Data analysis

Probability density plots were made by counting the total number of visits made by a given test group to each square millimetre of the assay plate and dividing by the total number of sample intervals. Instantaneous speed was calculated by measuring the displacement of the centroid of the worm once per second. We quantified chemotaxis performance by computing the percentage of time that the centroid of the animal was within a circular region (10-mm diameter) centred at the peak of the gradient. The results presented here remained qualitatively the same when the diameter of the peak region used to calculate chemotaxis performance was increased to 20 mm (data not shown). All averages are reported  $\pm 95$  per cent confidence intervals.

Received 7 November 2000; accepted 25 January 2001.

1. Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7**, 729–742 (1991).
2. Ward, S., Thompson, N., White, J. & Brenner, S. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* **160**, 313–338 (1975).
3. White, J., Southgate, E., Thompson, J. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
4. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94**, 3384–3387 (1997).
5. Hobert, O., Tessmar, K. & Ruvkun, G. The *Caenorhabditis elegans* *lim-6* LIM homeobox gene regulates neurite outgrowth and function of particular GABAergic neurons. *Development* **126**, 1547–1562 (1999).
6. Pierce-Shimomura, J. T., Morse, T. M. & Lockery, S. R. The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *J. Neurosci.* **19**, 9557–9569 (1999).
7. Ward, S. Chemotaxis by the nematode *Caenorhabditis elegans*: identification of attractants and analysis of the response by use of mutants. *Proc. Natl Acad. Sci. USA* **70**, 817–821 (1973).

8. Goodman, M. B., Hall, D. H., Avery, L. & Lockery, S. R. Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. *Neuron* **20**, 763–772 (1998).
9. Troemel, E. R., Sagasti, A. & Bargmann, C. I. Lateral signaling mediated by axon contact and calcium entry regulates asymmetric odorant receptor expression in *C. elegans*. *Cell* **99**, 387–398 (1999).
10. Bargmann, C. I., Thomas, J. H. & Horvitz, H. R. Chemosensory cell function in the behavior and development of *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **55**, 529–538 (1990).
11. Bargmann, C. I. & Horvitz, H. R. Control of larval development by chemosensory neurons in *Caenorhabditis elegans*. *Science* **251**, 1243–1246 (1991).
12. Kaplan, J. M. & Horvitz, H. R. A dual mechanosensory and chemosensory neuron in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 2227–2231 (1993).
13. Dusenbery, D. B. Analysis of chemotaxis in the nematode by counter current separation. *J. Exp. Zool.* **188**, 41–47 (1974).
14. Bargmann, C. I., Hartwig, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74**, 515–527 (1993).
15. Wes, P. D. & Bargmann, C. I. *C. elegans* odour discrimination requires asymmetric diversity in olfactory neurons. *Nature* **410**, 698–701 (2001).
16. Mori, K. & Yoshihara, Y. Molecular recognition and olfactory processing in the mammalian olfactory system. *Prog. Neurobiol.* **45**, 585–619 (1995).
17. Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1991).
18. Bargmann, C. I. & Avery, L. Laser killing of cells in *Caenorhabditis elegans*. *Methods Cell Biol.* **48**, 225–250 (1995).
19. Sengupta, P., Chou, J. H. & Bargmann, C. I. *odr-10* encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell* **84**, 899–909 (1996).
20. Lockery, S. R. & Goodman, M. B. Tight-seal whole-cell patch clamping of *C. elegans* neurons. *Methods Enzymol.* **295**, 201–217 (1998).

## Acknowledgements

We thank C. Bargmann and P. Wes for discussions and sharing unpublished data; O. Hobert for suggesting study of the *lim-6* mutant; J. H. Thomas for technical instruction; O. Hobert, D. Garbers and The *C. elegans* Genetics Center for strains; M. Moravec and J. Cervantes for technical assistance; and J. Eisen, M. Goodman, T. Morse and M. Westerfield for discussion. This work was supported by the National Science Foundation; the National Institute of Mental Health; the National Heart, Lung, and Blood Institute; the Office of Naval Research; The Sloan Foundation; and The Searle Scholars Program.

Correspondence and requests for materials should be addressed to S.R.L. (e-mail: shawn@lox.uoregon.edu).

## *C. elegans* odour discrimination requires asymmetric diversity in olfactory neurons

Paul D. Wes & Cornelia I. Bargmann

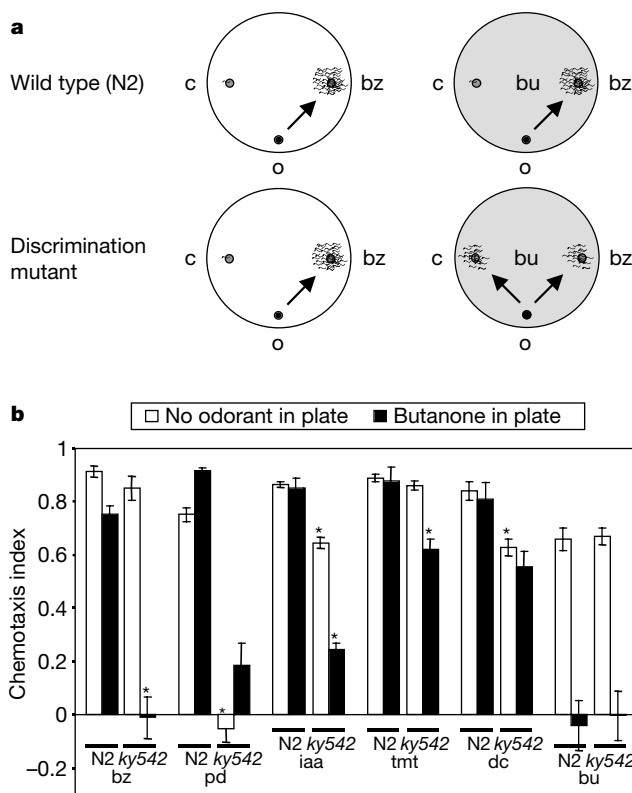
Howard Hughes Medical Institute, Programs in Developmental Biology, Neuroscience, and Genetics, Department of Anatomy and Department of Biochemistry and Biophysics, The University of California, San Francisco, California 94143-0452, USA

*Caenorhabditis elegans* senses at least five attractive odours with a single pair of olfactory neurons, AWC, but can distinguish among these odours in behavioural assays<sup>1</sup>. The two AWC neurons are structurally and functionally similar, but the G-protein-coupled receptor STR-2 is randomly expressed in either the left or the right AWC neuron, never in both<sup>2</sup>. Here we describe the isolation of a mutant, *ky542*, with specific defects in odour discrimination and odour chemotaxis. *ky542* is an allele of *nsy-1*, a neuronal symmetry, or *Nsy*, mutant in which STR-2 is expressed in both AWC neurons<sup>2</sup>. Other *Nsy* mutants exhibit discrimination and olfactory defects like those of *nsy-1* mutants. Laser ablation of the AWC neuron that does not express STR-2 (AWC<sup>OFF</sup>) recapitulates the behavioural phenotype of *Nsy* mutants, whereas laser ablation of the STR-2-expressing AWC neuron (AWC<sup>ON</sup>) causes different chemotaxis defects. We propose that odour discrimination can be achieved by segregating the detection of different odours

into distinct olfactory neurons or into unique combinations of olfactory neurons.

*Caenorhabditis elegans*, like other animals, detects odours with G-protein-coupled receptors such as the high-affinity diacetyl receptor encoded by the *odr-10* gene<sup>3</sup>. Individual chemosensory neurons express several receptor genes<sup>4,5</sup>, potentially allowing them to detect many odours but presenting a challenge for discriminating among odours. Nevertheless, *C. elegans* does exhibit the ability to distinguish between many odours in behavioural assays: a high uniform concentration of an odour will block chemotaxis toward a point source of that odour but not toward point sources of other odours<sup>1,6</sup>. The five odours sensed by the two AWC neurons— butanone, benzaldehyde, 2,3-pentanedione, isoamyl alcohol and 2,4,5-trimethylthiazole—use a common cyclic GMP signalling pathway, comprising the G<sub>α</sub> subunit ODR-3, the transmembrane guanylyl cyclases ODR-1 and DAF-11, and the cyclic-nucleotide-gated channel TAX-2/TAX-4 (refs 1, 6–10). Overexpression of ODR-1 perturbs discrimination between butanone and benzaldehyde through mechanisms that are not understood<sup>6</sup>.

To gain insight into the mechanism of discrimination, we performed a genetic screen for loss-of-function mutants that cannot detect a point source of benzaldehyde in a high uniform concentration of butanone (Fig. 1a; and Methods). The most striking defect was observed in the mutant *ky542*, which exhibited a complete loss of benzaldehyde chemotaxis in the presence of a



**Figure 1** The *ky542* mutant has defects in olfaction and olfactory discrimination.

**a**, Animals are scored for chemotaxis to a point source of benzaldehyde (bz) in the absence (chemotaxis assay, left) or presence (discrimination assay, right) of a uniform field of butanone (bu). Discrimination mutants fail to chemotax to benzaldehyde in the presence of butanone. c, control; o, origin. **b**, Population chemotaxis assays in the absence (open bars) or presence (filled bars) of butanone. Asterisks indicate statistically significant ( $P < 0.001$ ) defects in *ky542* mutants. Genotypes and odors are listed below each assay. N2, wild type; bu, 1:1,000 butanone; bz, 1:200 benzaldehyde; dc, 1:1,000 diacetyl; iaa, 1:100 isoamyl alcohol; pd, 1:10,000 2,3-pentanedione; tmt, 1:1,000 2,4,5-trimethylthiazole. The abbreviations are the same for all figures.

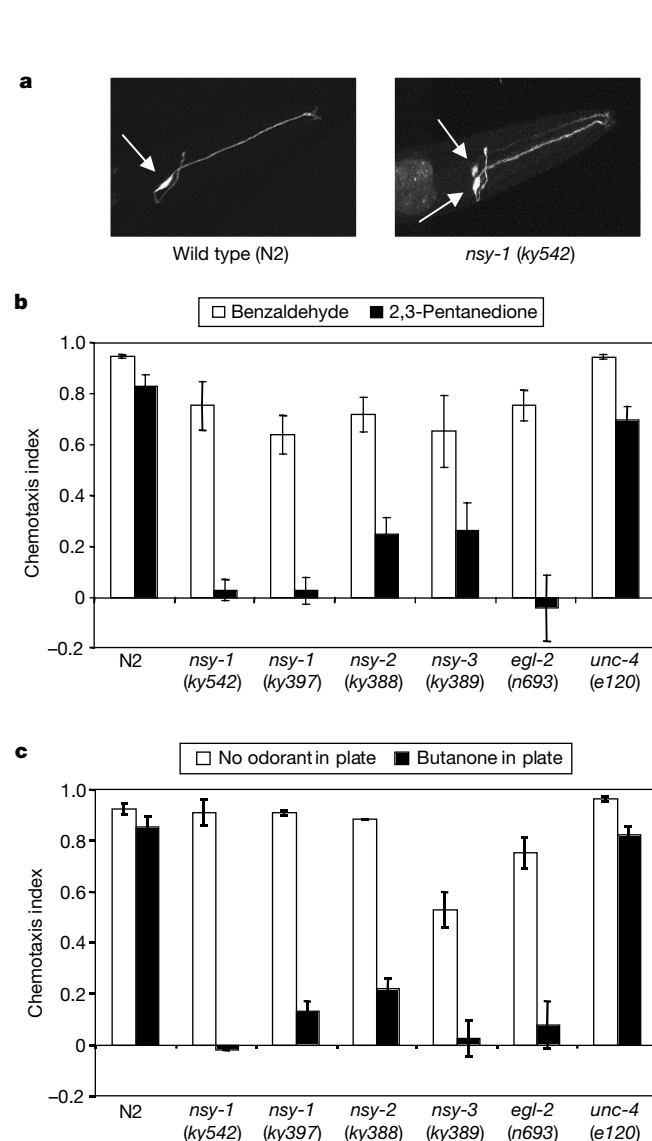
uniform field of butanone (Fig. 1b). In addition, the response of *ky542* mutants to isoamyl alcohol and 2,4,5-trimethylthiazole was moderately reduced in the presence of butanone (Fig. 1b). *ky542* mutants did not exhibit any discrimination defects in the presence of high uniform concentrations of benzaldehyde, 2,3-pentanedione, isoamyl alcohol and diacetyl (data not shown).

In addition to their defect in odour discrimination, *ky542* mutants did not detect 2,3-pentanedione in chemotaxis assays and were mildly defective in the detection of isoamyl alcohol and diacetyl (Fig. 1b). Diacetyl is sensed primarily by AWA, but also to some extent by AWC<sup>1</sup>. *ky542* mutants had normal detection of benzaldehyde, butanone and 2,4,5-trimethylthiazole. Thus, some AWC-mediated chemotaxis responses are perturbed in *ky542* mutants, whereas other responses that use the same signalling pathway are normal.

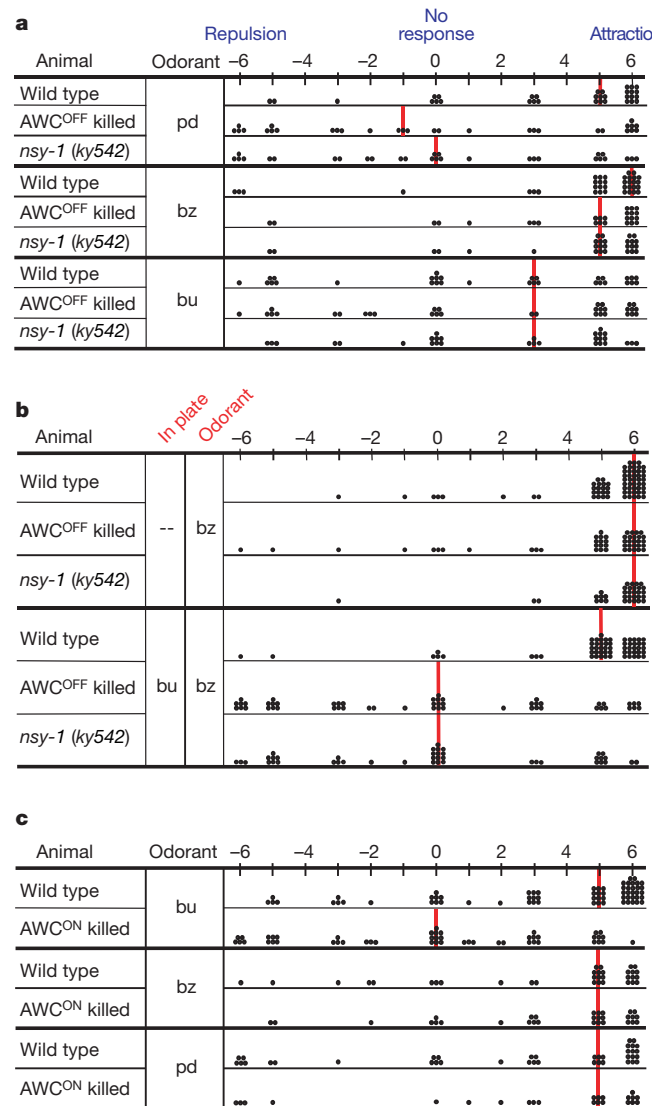
*ky542* was recessive and mapped to a 2 cM region on chromo-

some II. This region includes the neuronal symmetry gene, *nsy-1* (A. Sagasti, personal communication). Wild-type animals express *str-2::GFP* (green fluorescent protein) randomly in either the left or the right AWC neuron<sup>2</sup>. The *nsy-1* allele, *nsy-1(ky397)*, was isolated in a screen for mutants that express *str-2::GFP* in both AWC neurons<sup>2</sup>. *ky542* failed to complement the *str-2::GFP* expression phenotype of *nsy-1(ky397)*. Like *nsy-1(ky397)* animals<sup>2</sup>, 99% of *ky542* homozygous animals expressed *str-2::GFP* in both AWC neurons ( $n = 100$ ; Fig. 2a). Moreover, like *nsy-1(ky542)*, *nsy-1(ky397)* animals were defective in 2,3-pentanedione chemotaxis (Fig. 2b) and butanone/benzaldehyde discrimination (Fig. 2c). We conclude that *ky542* is an allele of *nsy-1*.

The *Nsy* phenotype can be caused by loss-of-function mutations in calcium channels and Ca<sup>2+</sup>/calmodulin-dependent kinase II,



**Figure 2** *Nsy* mutants have defects in olfaction and olfactory discrimination. **a**, In 100% of wild-type animals *str-2::GFP* is expressed in only one AWC neuron. In 99% of *nsy-1(ky542)* mutants, *str-2::GFP* is expressed in both AWC neurons. Arrows indicate AWC cell bodies; anterior is at right. **b**, **c**, The *Nsy* mutants *nsy-1(ky542)*, *nsy-1(ky397)*, *nsy-2(ky388)*, *nsy-3(ky389)* and *egl-2(n693)* did not chemotax toward 1:10,000 2,3-pentanedione (**b**, filled bars) or toward 1:200 benzaldehyde in the presence of butanone (**c**, filled bars), but did chemotax toward benzaldehyde in the absence of butanone (**b**, **c**, open bars). The slightly uncoordinated mutant *unc-4* behaved like wild-type animals in these assays.



**Figure 3** AWC<sup>OFF</sup> and AWC<sup>ON</sup> have distinct functions in olfaction and olfactory discrimination. Ablated animals were subjected to single-animal assays, along with unoperated siblings ('wild-type') and *nsy-1(ky542)* mutants. Each dot represents one single-animal chemotaxis assay; vertical lines represent the median response. **a**, AWC<sup>OFF</sup>-ablated animals ('AWC<sup>OFF</sup> killed',  $n = 18$  animals) failed to detect 2,3-pentanedione ( $P = 0.006$ ), but detected butanone and benzaldehyde. **b**, AWC<sup>OFF</sup>-ablated ( $n = 24$ ) and *nsy-1* animals exhibited a discrimination defect as they did not detect benzaldehyde in the presence of butanone, but responded normally in its absence ( $P < 0.001$ ). **c**, AWC<sup>ON</sup>-ablated animals ( $n = 41$ ) did not chemotaxis toward butanone ( $P < 0.001$ ), but responded normally to benzaldehyde and 2,3-pentanedione.



gain-of-function mutations in the eag-type potassium channel gene *egl-2*, and mutations in several uncloned *nsy* genes<sup>2,12</sup>. Odour discrimination was examined in all *Nsy* mutants that were not too uncoordinated to chemotax. *nsy-2*, *nsy-3* and *egl-2(gf)* mutants were defective both in 2,3-pentanedione chemotaxis (Fig. 2b) and in butanone/benzaldehyde discrimination (Fig. 2c).

All of the *Nsy* mutants share a common defect in *str-2::GFP* expression: the loss of the AWC<sup>OFF</sup> cell and the generation of two AWC<sup>ON</sup> cells. Therefore, one possible explanation for the behavioural phenotypes of *Nsy* mutants is that AWC<sup>OFF</sup> is required for 2,3-pentanedione detection and butanone/benzaldehyde discrimination. To test this possibility directly, we killed AWC<sup>OFF</sup> in wild-type animals by laser microsurgery, leaving AWC<sup>ON</sup> intact. Ablations were performed at a developmental stage when the AWC<sup>ON</sup> and AWC<sup>OFF</sup> fates are no longer altered by killing either cell<sup>2</sup>.

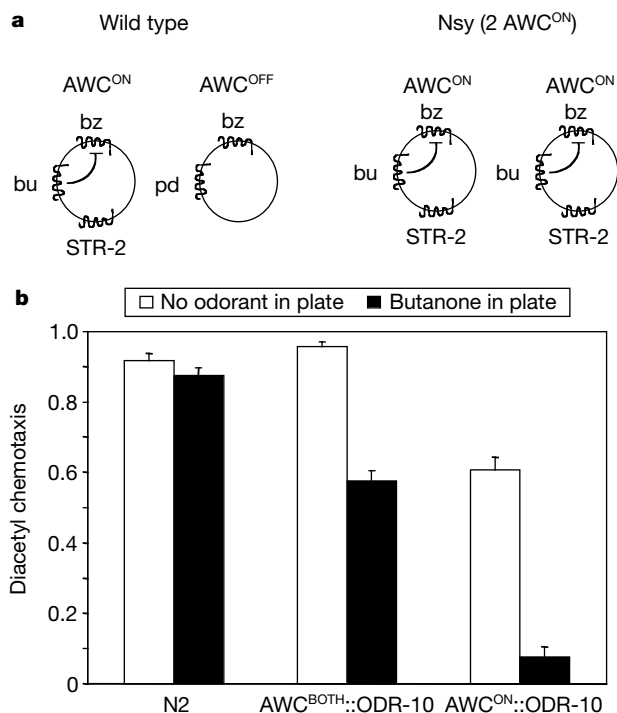
AWC<sup>OFF</sup> ablations recapitulated both of the behavioural defects of *Nsy* mutants: AWC<sup>OFF</sup>-ablated animals did not detect 2,3-pentanedione (Fig. 3a) and did not discriminate benzaldehyde from butanone (Fig. 3b). Chemotaxis to butanone and benzaldehyde was normal in AWC<sup>OFF</sup>-ablated animals (Fig. 3a). Ablation of AWC<sup>ON</sup> abolished chemotaxis to butanone but did not affect chemotaxis to benzaldehyde or 2,3-pentanedione (Fig. 3c). Killing AWC<sup>ON</sup> did not cause defects in butanone/benzaldehyde discrimination (data not shown). Thus, AWC<sup>OFF</sup> and AWC<sup>ON</sup> are each required for the responses to a specific odour, 2,3-pentanedione

and butanone, respectively, but share responses to another odour, benzaldehyde. Furthermore, AWC<sup>OFF</sup>, but not AWC<sup>ON</sup>, is required for discrimination between butanone and benzaldehyde.

To explain why the loss of AWC<sup>OFF</sup> produces a discrimination defect, we propose that a high concentration of butanone can attenuate or block olfactory signalling to benzaldehyde in the cell that detects butanone (Fig. 4a). Because only one AWC neuron detects butanone (AWC<sup>ON</sup>) but both AWC neurons detect benzaldehyde in wild-type animals, butanone can attenuate benzaldehyde signalling only in AWC<sup>ON</sup>. In the presence of butanone, therefore, AWC<sup>OFF</sup> continues to mediate normal chemotaxis toward benzaldehyde. In *Nsy* mutants, both AWC neurons detect butanone, and therefore butanone attenuates benzaldehyde signalling in both neurons, abolishing benzaldehyde chemotaxis. In AWC<sup>OFF</sup>-ablated animals, the one surviving AWC<sup>ON</sup> neuron senses both butanone and benzaldehyde, so butanone abolishes benzaldehyde chemotaxis in this context as well.

To test this model directly, we expressed the diacetyl receptor ODR-10 in both AWC neurons (AWC<sup>BOTH</sup>::ODR-10) or solely in the AWC<sup>ON</sup> neuron (AWC<sup>ON</sup>::ODR-10). The *odr-10* gene is normally expressed in the AWA olfactory neurons, but not in the AWC neurons<sup>3</sup>. In these experiments, we inactivated endogenous ODR-10 function in AWA using an *odr-7* or an *odr-10* mutation<sup>3,13</sup>. AWC<sup>BOTH</sup>::ODR-10 restored full diacetyl chemotaxis to mutant animals, whereas AWC<sup>ON</sup>::ODR-10 supported partial chemotaxis (Fig. 4b). In the presence of butanone, chemotaxis towards diacetyl was abolished in AWC<sup>ON</sup>::ODR-10, but not in AWC<sup>BOTH</sup>::ODR-10 (Fig. 4b). Butanone had no effect on diacetyl chemotaxis in a wild-type background. These results are consistent with butanone attenuating diacetyl signalling only in the AWC<sup>ON</sup> cell.

Odour discrimination increases the ability of an animal to sample salient features of a complex environment by ignoring pervasive odours. Our results indicate that the two AWC olfactory neurons are functionally distinct, and that loss of this diversity in *Nsy* mutants leads to defects in chemotaxis and odour discrimination. A similar functional distinction between otherwise symmetric chemosensory neurons was observed<sup>14</sup> in the ASE taste neurons and found subsequently to affect taste discrimination. Like the *Nsy* mutants, animals that overexpress the guanylyl cyclase ODR-1 do not effectively discriminate between butanone and benzaldehyde<sup>6</sup>. A neuronal symmetry defect may be involved in this phenotype, as 50% of these animals also express *str-2::GFP* in both AWC neurons (N. L'Etoile, personal communication). Segregation of receptor expression to distinct neurons cannot account for all the odour discrimination observed in *C. elegans*, however, as the two AWC neurons can discriminate at least five odours, and some forms of odour discrimination in AWC are normal in *nsy-1* mutants. Other forms of odour discrimination may be achieved within a single neuron, if different odour receptors interact with distinct regulators or effector molecules, or if different odour receptors are sequestered into discrete signalling complexes. Thus, the segregation of receptors into different neurons may be one of many mechanisms that shape and regulate olfactory responses.



**Figure 4** A model for the behavioural defects observed in *Nsy* mutants. **a**, In wild-type, both AWC neurons recognize benzaldehyde, AWC<sup>ON</sup> recognizes butanone and AWC<sup>OFF</sup> recognizes 2,3-pentanedione. Butanone signalling attenuates benzaldehyde signalling in AWC<sup>ON</sup> (arrow). In *Nsy* mutants (right), the AWC<sup>OFF</sup> neuron is transformed into AWC<sup>ON</sup>, 2,3-pentanedione chemotaxis is lost, and butanone attenuates benzaldehyde signalling in both neurons. **b**, Butanone attenuates ODR-10 signalling in AWC<sup>ON</sup>. AWC<sup>ON</sup>::ODR-10 animals express *str-2::ODR-10* in AWC<sup>ON</sup> in an *odr-10* mutant background. AWC<sup>BOTH</sup>::ODR-10 animals express *odr-3::ODR-10* in both AWC neurons<sup>9</sup> in an *odr-7* mutant background. *odr-3* is expressed in AWA, but the *odr-7* mutation inactivates AWA<sup>9,13</sup>. Diacetyl chemotaxis is partly (AWC<sup>ON</sup>::ODR-10) or entirely (AWC<sup>BOTH</sup>::ODR-10) rescued (open bars). When butanone is present (filled bars), diacetyl chemotaxis is abolished in AWC<sup>ON</sup>::ODR-10 animals, reduced to the partly rescued level in AWC<sup>BOTH</sup>::ODR-10 animals and unaffected in wild-type animals.

## Methods

### Chemotaxis assays

Population and single-animal chemotaxis assays were done as described<sup>1,15</sup>. Each data point for population assays represents the mean of a minimum of six assays performed on two different days (up to 44 assays performed on 19 days); error bars indicate the standard error of the mean. The statistical significance of population and single-animal chemotaxis assays was determined using the *t*-test and Mann–Whitney rank sum test, respectively. For discrimination assays, butanone was added to a final concentration of 1.2  $\mu$ l per 10-ml plate, and mixed with the liquid agar once it had cooled to 55 °C.

### Isolation of mutants

Wild-type N2 animals were mutagenized with ethylmethanesulphonate using standard protocols<sup>16</sup>. Animals from the F<sub>2</sub> generation (representing ~9,000 mutagenized genomes) were washed three times in S Basal and once in distilled water, divided into batches of

roughly 300 animals, and given a choice between an optimal concentration of benzaldehyde (1:200 dilution in ethanol) and a lower concentration of diacetyl (1:10,000 dilution in ethanol) in the presence of a uniform field of butanone (1.2  $\mu$ l per 10-ml plate). Under these conditions more than 95% of wild-type animals prefer benzaldehyde. Animals that accumulated at the diacetyl source were removed and retested under the same conditions to repeat the enrichment. Animals that preferred diacetyl three times were isolated, and their  $F_2$  broods were given a choice between benzaldehyde and diacetyl in the absence of a uniform concentration of butanone. Mutants that could chemotax to benzaldehyde under these conditions were saved. Twenty-seven mutants exhibited discrimination defects that could also be replicated without the diacetyl counterattractant. Mutants were backcrossed twice to wild-type animals.

### Genetic mapping of *ky542*

We mapped *ky542* to chromosome II by observing segregation of the discrimination phenotype away from the dominant marker *sqt-1(scl)* (7/7 isolates). Subsequent mapping was performed by following segregation of the discrimination phenotype with single-nucleotide polymorphisms (SNPs) between the wild-type N2 and CB4856 strains.  $F_2$  progeny of *ky542*  $\times$  CB4856 crosses were isolated, and populations were generated from each isolate. Each population was tested for butanone/benzaldehyde discrimination. Populations that were homozygous mutant and those that were homozygous wild type were retained, whereas populations that appeared to be heterozygous were discarded. We isolated DNA from each population, and scored SNPs by polymerase chain reaction amplification followed by restriction-enzyme digestion. Using 33 populations, we found that *ky542* mapped between SNPs located on cosmid C01F1 (chromosome II, position -4.5) and cosmid C34F1 (chromosome II, position -2.5).

### Laser ablations

AWC neurons were ablated in a wild-type strain that contained an integrated *str-2::GFP* reporter (*kyIs140*) at the L1 or L2 larval stage<sup>17</sup>. The AWC neuron was identified by its characteristic position or by the use of the *str-2::GFP* marker, and then laser irradiated. Ablation was confirmed for AWC<sup>ON</sup>-ablated animals by looking for *str-2::GFP* expression after all assays had been performed. Single-animal assays were performed on gravid adults as early as the second day after ablation and as late as the fourth day. We assayed the same animals on two or three consecutive days. As many as three consecutive olfactory assays were performed in a single day. For discrimination assays, in which animals were challenged with the same attractant in the presence and absence of saturating odour, animals were allowed to recover between tests for 1 h on a fresh plate with no odours. The order of the assays was randomized on different days. Single-animal assay plates were poured 1 day before the assays and allowed to air dry for 1 h before the assays.

Received 6 November 2000; accepted 11 January 2001.

- Bargmann, C. I., Hartwig, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74**, 515–527 (1993).
- Troemel, E. R., Sagasti, A. & Bargmann, C. I. Lateral signaling mediated by axon contact and calcium entry regulates asymmetric odorant receptor expression in *C. elegans*. *Cell* **99**, 387–398 (1999).
- Sengupta, P., Chou, J. H. & Bargmann, C. I. *odr-10* encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell* **84**, 899–909 (1996).
- Troemel, E. R., Chou, J. H., Dwyer, N. D., Colbert, H. A. & Bargmann, C. I. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell* **83**, 207–218 (1995).
- Troemel, E. R. *Chemosensory Receptors in Caenorhabditis elegans*. Ph.D. Thesis, Univ. California (1999).
- L'Etoile, N. D. & Bargmann, C. I. Olfaction and odor discrimination are mediated by the *C. elegans* guanylyl cyclase ODR-1. *Neuron* **25**, 575–586 (2000).
- Komatsu, H., Mori, I., Rhee, J.-S., Akaike, N. & Ohshima, Y. Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. *Neuron* **17**, 707–718 (1996).
- Coburn, C. M. & Bargmann, C. I. A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. *Neuron* **17**, 695–706 (1996).
- Roayaie, K., Crump, J. G., Sagasti, A. & Bargmann, C. I. The G alpha protein ODR-3 mediates olfactory and nociceptive function and controls cilium morphogenesis in *C. elegans* olfactory neurons. *Neuron* **20**, 55–67 (1998).
- Birnby, D. A., Link, E. M., Vowles, J. J., Tian, H., Colacurcio, P. L. & Thomas, J. H. A transmembrane guanylyl cyclase (DAF-11) and Hsp90 (DAF-21) regulate a common set of chemosensory behaviors in *Caenorhabditis elegans*. *Genetics* **155**, 85–104 (2000).
- Chou, J. H., Bargmann, C. I. & Sengupta, P. The *Caenorhabditis elegans odr-2* gene encodes a novel Ly-6-related protein required for olfaction. *Genetics* **157**, 211–224 (2001).
- Weinschenker, D., Wei, A., Salkoff, L. & Thomas, J. H. Block of an ether-a-go-go-like  $K^+$  channel by imipramine rescues *egl-2* excitation defects in *Caenorhabditis elegans*. *J. Neurosci.* **19**, 9831–9840 (1999).
- Sengupta, P., Colbert, H. A. & Bargmann, C. I. The *C. elegans* gene *odr-7* encodes an olfactory-specific member of the nuclear receptor superfamily. *Cell* **79**, 971–980 (1994).
- Pierce-Shimomura, J. T., Faumont, S., Gaston, M. R., Pearson, B. J. & Lockery, S. R. The homeobox gene *lim-6* is required for distinct chemosensory representations in *C. elegans*. *Nature* **410**, 694–698 (2001).
- Troemel, E. R., Kimmel, B. E. & Bargmann, C. I. Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans*. *Cell* **91**, 161–169 (1997).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7**, 729–742 (1991).

### Acknowledgements

We thank A. Sagasti and N. L'Etoile for discussions and insights into AWC asymmetry; H. Nguyen for technical assistance; S. Wicks and R. Plasterk for sharing unpublished data; C. Adler and S. Kirch for advice on microscopy; and A. Sagasti, S. Shaham, N. L'Etoile, J. Gray and C. Adler for comments on the manuscript. P.D.W. was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. C.I.B. is an Investigator of the Howard Hughes Medical Institute. This work was supported by a grant from the National Institutes of Health.

Correspondence and requests for materials should be addressed to C.I.B. (e-mail: cori@itsa.ucsf.edu).

## Bone marrow cells regenerate infarcted myocardium

Donald Orlie<sup>†</sup>, Jan Kajstura<sup>\*</sup>, Stefano Chimenti<sup>\*</sup>, Igor Jakoniuk<sup>\*</sup>, Stacie M. Anderson<sup>†</sup>, Baosheng Li<sup>\*</sup>, James Pickel<sup>‡</sup>, Ronald McKay<sup>‡</sup>, Bernardo Nadal-Ginard<sup>\*</sup>, David M. Bodine<sup>†</sup>, Annarosa Leri<sup>\*</sup> & Piero Anversa<sup>\*</sup>

<sup>\*</sup> Department of Medicine, New York Medical College, Valhalla, New York 10595, USA

<sup>†</sup> Hematopoiesis Section, Genetics and Molecular Biology Branch, NHGRI, and

<sup>‡</sup> Laboratory of Molecular Biology, NINDS, NIH, Bethesda, Maryland 20892, USA

Myocardial infarction leads to loss of tissue and impairment of cardiac performance. The remaining myocytes are unable to reconstitute the necrotic tissue, and the post-infarcted heart deteriorates with time<sup>1</sup>. Injury to a target organ is sensed by distant stem cells, which migrate to the site of damage and undergo alternate stem cell differentiation<sup>2–5</sup>; these events promote structural and functional repair<sup>6–8</sup>. This high degree of stem cell plasticity prompted us to test whether dead myocardium could be restored by transplanting bone marrow cells in infarcted mice. We sorted lineage-negative ( $Lin^-$ ) bone marrow cells from transgenic mice expressing enhanced green fluorescent protein<sup>9</sup> by fluorescence-activated cell sorting on the basis of *c-kit* expression<sup>10</sup>. Shortly after coronary ligation,  $Lin^- c-kit^{POS}$  cells were injected in the contracting wall bordering the infarct. Here we report that newly formed myocardium occupied 68% of the infarcted portion of the ventricle 9 days after transplanting the bone marrow cells. The developing tissue comprised proliferating myocytes and vascular structures. Our studies indicate that locally delivered bone marrow cells can generate *de novo* myocardium, ameliorating the outcome of coronary artery disease.

Injection of male  $Lin^- c-kit^{POS}$  bone marrow cells (see Supplementary Information) in the peri-infarcted left ventricle of female mice resulted in myocardial regeneration. Repair was obtained in 12 out of 30 mice (40%). Failure to reconstitute infarcts was attributed to the difficulty of transplanting cells into tissue contracting at 600 beats per minute. However, an immunological reaction to the histocompatibility antigen on the Y chromosome of the donor bone marrow cells could account for the lack of repair in some of the female recipients. Closely packed myocytes occupied  $68 \pm 11\%$  of the infarcted region and extended from the anterior to the posterior aspect of the ventricle (Fig. 1a–d). The fraction of endocardial and epicardial circumference delimiting the infarcted area<sup>1,11</sup> did not differ in untreated mice,  $78 \pm 18\%$  ( $n = 8$ ), or in mice treated with  $Lin^- c-kit^{POS}$  cells,  $75 \pm 14\%$  ( $n = 12$ ), or  $Lin^- c-kit^{NEG}$  cells,  $75 \pm 15\%$  ( $n = 11$ ). New myocytes were not found in mice injected with  $Lin^- c-kit^{NEG}$  cells (Fig. 1e).

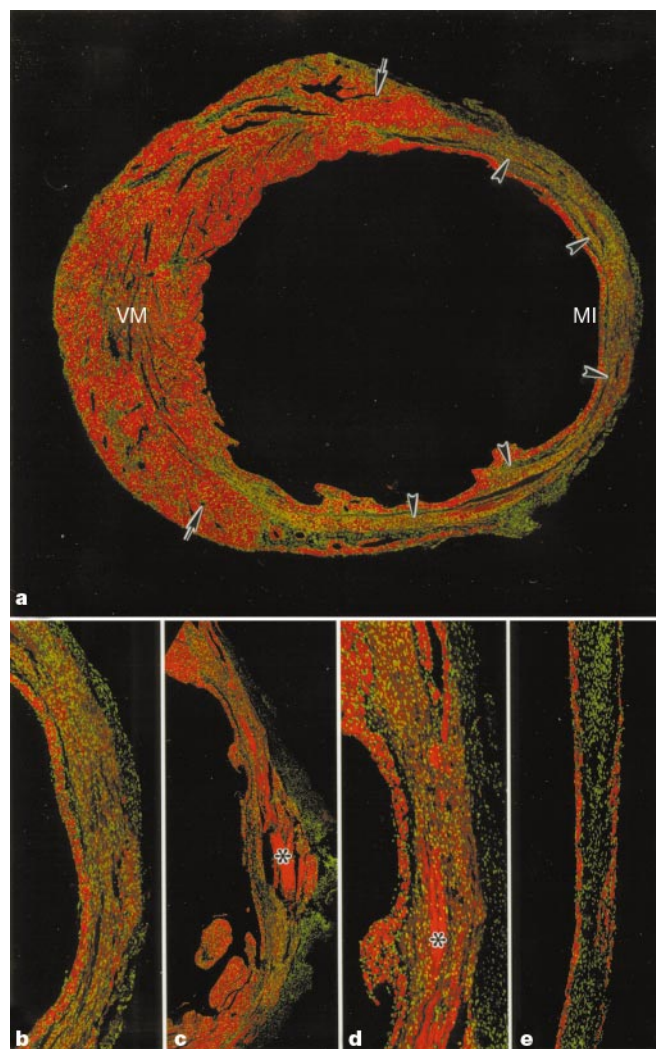
The origin of the cells in the forming myocardium was deter-



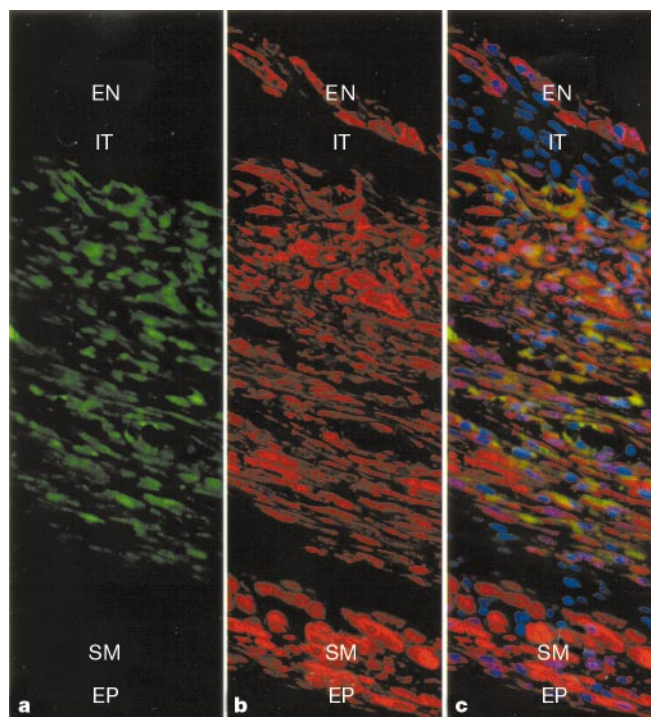
mined by the expression of enhanced green fluorescent protein (EGFP) (Fig. 2; see also Supplementary Information) and the presence of Y chromosome (Supplementary Information). EGFP was restricted to the cytoplasm, whereas Y chromosome was restricted to the nuclei of new cardiac cells. EGFP and Y chromosome were not detected in the surviving portion of the ventricle. EGFP expression was combined with the labelling of proteins specific for myocytes, endothelial cells and smooth muscle cells. This allowed us to identify each cardiac cell type, and to recognize endothelial and smooth muscle cells organized in coronary vessels (Fig. 3a–c; see also Supplementary Information). The percentage of new myocytes, endothelial cells and smooth muscle cells expressing EGFP was  $53 \pm 9\%$  ( $n = 7$ ),  $44 \pm 6\%$  ( $n = 7$ ) and  $49 \pm 7\%$  ( $n = 7$ ), respectively. These values were consistent with the fraction of transplanted  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  bone marrow cells that expressed EGFP,  $44 \pm 10\%$  ( $n = 6$ ). An average  $54 \pm 8\%$  ( $n = 6$ ) of myocytes, endothelial cells and smooth muscle cells expressed EGFP in the heart of donor transgenic mice.

To confirm that newly formed myocytes represented maturing

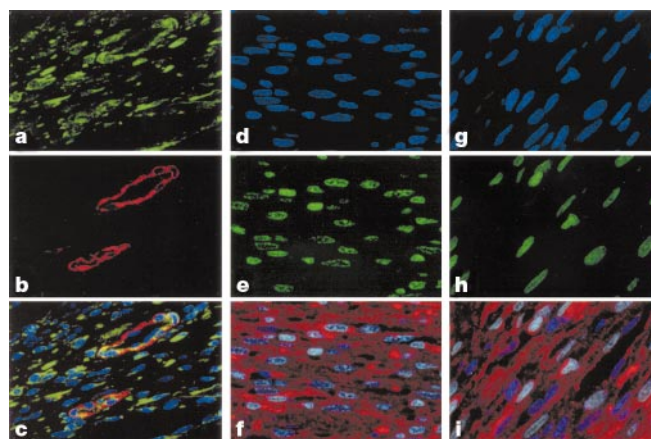
cells aiming at functional competence, we examined expression of the myocyte enhancer factor 2 (MEF2), the cardiac specific transcription factor GATA-4 and the early marker of myocyte development *Csx/Nkx2.5*. In the heart, MEF2 proteins are recruited by GATA-4 to activate synergistically the promoters of several cardiac genes, such as myosin light chain, troponin T, troponin I,  $\alpha$ -myosin heavy chain, desmin, atrial natriuretic factor and  $\alpha$ -actin<sup>12,13</sup>. *Csx/Nkx2.5* is a transcription factor restricted to the initial phases of



**Figure 1** Bone marrow cells and myocardial regeneration. **a**, Myocardial infarct (MI) injected with  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells from bone marrow (arrows). Arrowheads indicate regenerating myocardium; VM, viable myocardium; MI, myocardial infarct. **b**, Same MI at higher magnification. **c**, **d**, Low and high magnifications of MI injected with  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells. **e**, MI injected with  $\text{Lin}^- \text{c-kit}^{\text{NEG}}$  cells; only healing is apparent. Asterisk indicates necrotic myocytes. Red, cardiac myosin; green, propidium iodide labelling of nuclei. Original magnification,  $\times 12$  (**a**);  $\times 25$  (**c**)  $\times 50$  (**b**, **d**, **e**).



**Figure 2** Myocardial infarct injected with  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells; myocardium is regenerating from endocardium (EN) to epicardium (EP). **a**, EGFP (green); **b**, cardiac myosin (red); **c**, combination of EGFP and myosin (red–green), and propidium-iodide-stained nuclei (blue). Infarcted tissue (IT) can be seen in the subendocardium; spared myocytes (SM) can be seen in the subepicardium. Original magnification,  $\times 250$  (**a–c**).



**Figure 3** Regenerating myocardium in myocardial infarct injected with  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells. **a**, EGFP (green); **b**, smooth muscle  $\alpha$ -actin in arterioles (red); **c**, combination of EGFP and smooth muscle  $\alpha$ -actin (yellow–red), and propidium iodide (PI)-stained nuclei (blue). **d–i**, MEF2 and *Csx/Nkx2.5* in cardiac myosin-positive cells. **d**, **g**, PI-stained nuclei (blue); **e**, **h**, MEF2 and *Csx/Nkx2.5* labelling (green); **f**, **i**, cardiac myosin (red), and combination of MEF2 or *Csx/Nkx2.5* with PI (bright fluorescence in nuclei). Original magnification,  $\times 300$  (**a–i**).



myocyte differentiation<sup>12</sup>. In the reconstituting heart, all nuclei of cells labelled with cardiac myosin expressed MEF2 (Fig. 3d–f) and GATA-4 (Supplementary Information), but only  $40 \pm 9\%$  expressed Csx/Nkx2.5 (Fig. 3g–i).

To characterize further the properties of these myocytes, we determined the expression of connexin 43. This protein is responsible for intercellular connections and electrical coupling through the generation of plasma-membrane channels between myocytes<sup>14,15</sup>; connexin 43 was apparent in the cell cytoplasm and at the surface of closely aligned differentiating cells (Fig. 4). These results were consistent with the expected functional competence of the heart muscle phenotype. In addition, myocytes at various stages of maturation were detected within the same and different bands (Fig. 5).

Ki67 is expressed in cycling cells in G1, S, G2 and early mitosis<sup>16</sup>, providing a quantitative estimate of the fraction of cells in the cell cycle at the time of observation. 5-Bromodeoxyuridine (BrdU) labelling identifies nuclei in S phase<sup>16,17</sup>; therefore, we injected BrdU for 4–5 days to assess cumulative cell division during active growth (Supplementary Information). Proliferation of myocytes was 93% ( $P < 0.001$ ) and 60% ( $P < 0.001$ ) higher than that of endothelial cells, and 225% ( $P < 0.001$ ) and 176% ( $P < 0.001$ ) higher than that of smooth muscle cells, when measured by BrdU and Ki67, respectively (BrdU: myocytes  $36 \pm 8\%$ ; endothelial cells  $19 \pm 5\%$ ; smooth muscle cells  $11 \pm 2\%$ ; Ki67: myocytes  $19 \pm 3\%$ ; endothelial cells  $12 \pm 3\%$ ; smooth muscle cells  $7 \pm 2\%$ ;  $n = 8$  in all cases). Dividing myocytes were small with partially aligned myofibrils, resembling late fetal/neonatal cells; 40–50% of the Ki67- or BrdU-positive cells expressed EGFP.

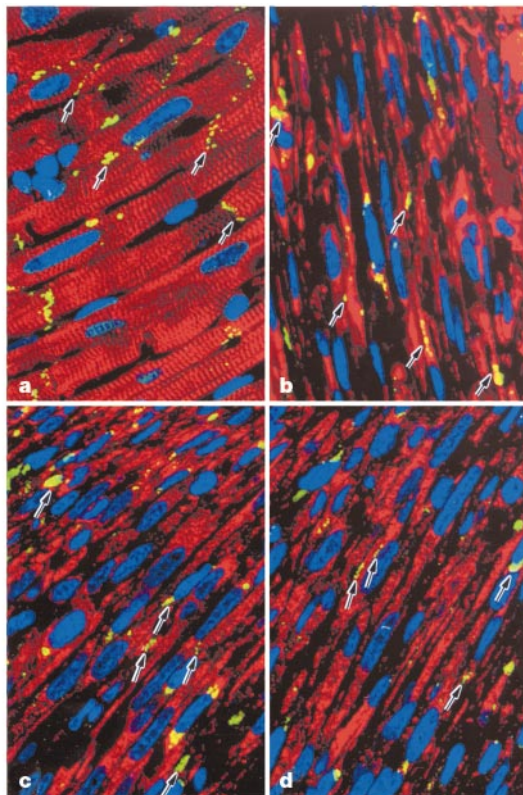
Cell differentiation caused a loss of *c-kit* surface receptors. We observed only two undifferentiated cells showing *c-kit* on the cell

membrane in the subendocardium of the infarcted wall. These *c-kit*-labelled cells were in proximity but not within the regenerating band. They expressed EGFP, confirming their origin from the transplant (Supplementary Information).

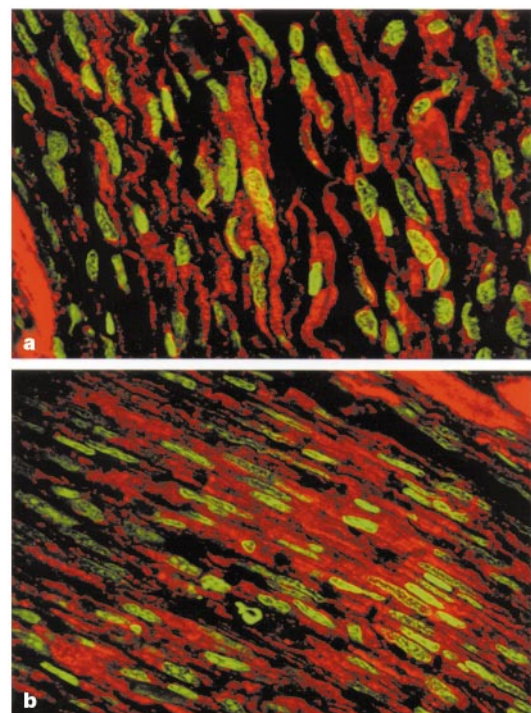
To determine whether developing myocytes derived from the  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells had an impact on function, we obtained haemodynamic parameters before death. Results from infarcted mice non-injected or injected with  $\text{Lin}^- \text{c-kit}^{\text{NEG}}$  cells were combined. In comparison with sham-operated mice, the infarcted groups exhibited indices of cardiac failure (Fig. 6a). In mice treated with  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells, left ventricular (LV) end-diastolic pressure (LVEDP) was 36% lower, and developed pressure (LVDP), LV + dP/dt and LV – dP/dt were 32%, 40% and 41% higher, respectively.

Locally transplanted  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  bone marrow cells have a high capacity for cardiac tissue differentiation. Here, they led to the formation of new myocytes, endothelial cells and smooth muscle cells generating *de novo* myocardium, inclusive of coronary arteries, arterioles and capillaries. The partial repair of the infarcted heart implies that the transplanted cells responded to signals from the injured myocardium that promoted their migration, proliferation and differentiation within the necrotic area of the ventricular wall (Fig. 6b). These differentiating myocytes expressed nuclear and cytoplasmic proteins typical of cardiac tissue. The presence of connexin 43 points to cellular coupling and functional competence of the restored myocardium (Fig. 6b). With postnatal maturation, stem cell function was assumed previously to be restricted to cell lineages present in the organ from which they are derived. However, this limitation in stem cell differentiation potential has been challenged by studies showing that bone marrow and neural stem cells can produce many cell types<sup>4,5,18–20</sup>. We report, for the first time, that a subpopulation of primitive bone marrow cells regenerate myocardium *in vivo*, replacing dead tissue.

Haematopoietic stem cells (HSCs), neural-crest-derived melanoblasts and primordial germ cells express *c-kit* on their cell membrane. These primitive cells migrate during fetal development,



**Figure 4** Myocardial repair and connexin 43. **a**, Border zone; **b–d**, regenerating myocardium. Shown are connexin 43 (yellow–green; arrows indicate contacts between myocytes) and  $\alpha$ -sarcomeric actin (red), and PI-stained nuclei (blue). Original magnification,  $\times 500$  (**a**),  $\times 800$  (**b–d**).



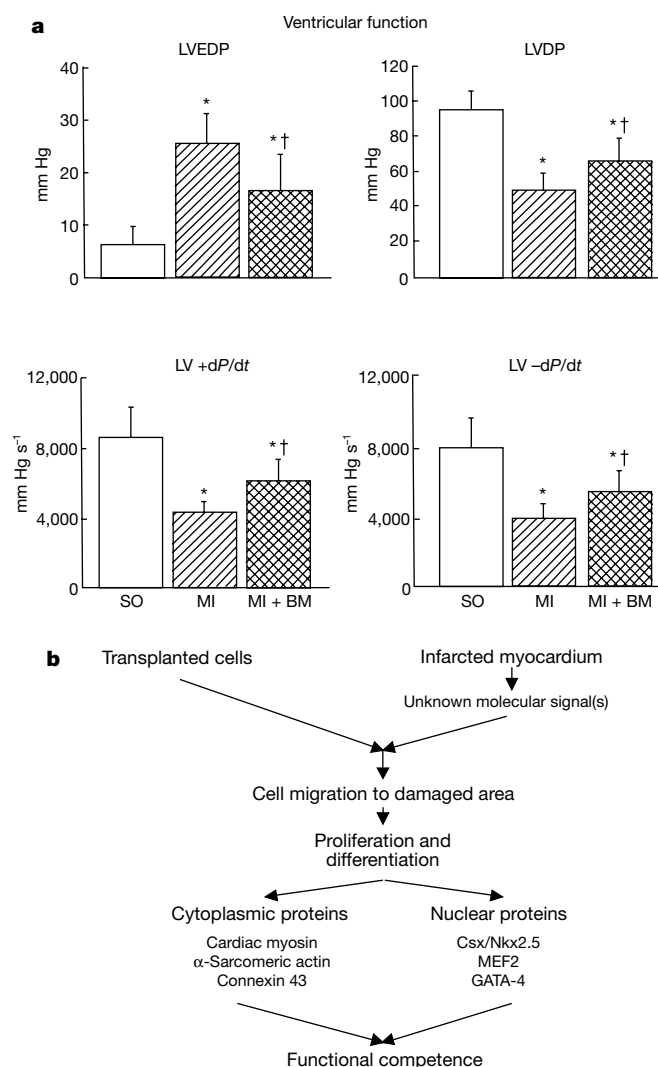
**Figure 5** Myocardial infarcts injected with  $\text{Lin}^- \text{c-kit}^{\text{POS}}$  cells: regenerating myocytes. Shown are cardiac myosin (red), and propidium-iodide-labelled nuclei (yellow–green). Original magnification,  $\times 1,000$  (**a**);  $\times 700$  (**b**).

homing to the yolk sac and liver. Both of these organs are positive for messenger RNA encoding stem cell factor (SCF), the ligand for *c-kit*<sup>21</sup>. It is thought that membrane-bound SCF mediates the migration of HSC and other primitive cells to their target organs<sup>22</sup>. The fetal and neonatal hearts are positive for SCF transcripts<sup>21</sup> and, although it is not clear whether adult heart cells generate SCF, the *c-kit*/SCF pathway might be the mechanism by which, in our conditions, transplanted Lin<sup>-</sup>*c-kit*<sup>POS</sup> cells migrated from the site of injection to the infarcted myocardium.

When a stem cell divides, two daughter cells are formed; these may maintain stem cell properties or become differentiating cells<sup>23</sup> that multiply much more rapidly than stem cells<sup>24</sup>. The Lin<sup>-</sup>*c-kit*<sup>POS</sup> cells in these transplants produced the three main cell types of the heart: myocytes constituted the predominant and most active growth component of the regenerating myocardium; endothelial and smooth muscle cells were fast growing but were smaller fractions of the developing tissue. Our observations are difficult

to compare with those obtained in the cryo-injured rat heart after injecting cultured myocytes derived from mesenchymal bone marrow cells<sup>25</sup>. Formation of myotubules *in vitro* was required for successful transplantation in that study<sup>25</sup>, which contrasts with our results. Cryo-injury has no human counterpart. It constitutes an unusual damage with an intact coronary circulation. This may be why only a few endothelial cells were possibly linked to the original culture system<sup>25</sup> and smooth muscle cells were not detected. Also at variance with our data is the fact that there was no replacement of damaged myocardium with functioning tissue.

Coronary heart disease accounts for 50% of all cardiovascular deaths and nearly 40% of the incidence of heart failure. The current findings have provided compelling evidence that our approach has relevant implications for human disease. Locally delivered primitive bone marrow cells promoted successful treatment of large myocardial infarcts after the completion of ischaemic cell death. This therapeutic intervention reduced the infarcted area and improved cardiac haemodynamics. Infarct size is a major determinant of morbidity and mortality, as massive infarcts affecting 40% or more of the left ventricle in patients are associated with intractable cardiogenic shock or the rapid development of congestive heart failure<sup>1</sup>. In the past, recovery of cardiac function has been fully dependent on the growth of the remaining non-infarcted portion of the ventricle. However, the hypertrophied infarcted heart undergoes progressive deterioration, leading to a dilated myopathy, terminal failure and death<sup>1</sup>. Transplanted Lin<sup>-</sup>*c-kit*<sup>POS</sup> bone marrow cells have the capability of regenerating acutely significant amounts of contracting myocardium. This new form of repair can improve the immediate and long-term outcome of ischaemic cardiomyopathy. □



**Figure 6** Postulated mechanism of myocardial regeneration and its effect on ventricular function. **a**, Effects of myocardial infarction (MI) on left ventricular end-diastolic pressure (LVEDP), developed pressure (LVDP), LV + dP/dt (rate of pressure rise) and LV - dP/dt (rate of pressure decay). Results are from sham-operated mice (SO,  $n = 11$ ), mice non-injected with Lin<sup>-</sup>*c-kit*<sup>POS</sup> cells (MI;  $n = 5$ ), mice injected with Lin<sup>-</sup>*c-kit*<sup>NEG</sup> cells ( $n = 6$ ), and mice injected with Lin<sup>-</sup>*c-kit*<sup>POS</sup> cells (MI+BM,  $n = 9$ ). Values are mean  $\pm$  s.d. \* $P < 0.05$  versus SO; † $P < 0.05$  versus MI. **b**, Proposed scheme for Lin<sup>-</sup>*c-kit*<sup>POS</sup> cell differentiation in cardiac muscle and functional implications.

## Methods

### Lin<sup>-</sup>*c-kit*<sup>POS</sup> cells

We collected bone marrow from the femurs and tibias of male transgenic mice expressing EGFP<sup>9</sup>. Cells were suspended in PBS containing 5% fetal calf serum (FCS) and incubated on ice with rat anti-mouse monoclonal antibodies specific for the following haematopoietic lineages: CD4 and CD8 (T lymphocytes), B-220 (B lymphocytes), Mac-1 (macrophages), GR-1 (granulocytes) (all Caltag Laboratories) and TER-119 (erythrocytes) (Pharmingen). Cells were then rinsed in PBS and incubated for 30 min with magnetic beads coated with goat anti-rat immunoglobulin (Pylisciences). Lineage-positive cells were removed by a biomagnet and the 10% remaining lineage-negative (Lin<sup>-</sup>) cells were stained with ACK-4-biotin (anti-*c-kit* monoclonal antibody). Cells were rinsed in PBS, stained with streptavidin-conjugated phycoerythrin (SA-PE) (Caltag) and sorted by FACS using a FACS Vantage instrument (Becton Dickinson). We excited EGFP and ACK-4-biotin-SA-PE at a wavelength of 488 nm. The Lin<sup>-</sup> cells were sorted as *c-kit*-positive (*c-kit*<sup>POS</sup>) and *c-kit*-negative (*c-kit*<sup>NEG</sup>) with a 1–2 log difference in staining intensity. The *c-kit*<sup>POS</sup> cells were suspended at a concentration of  $3 \times 10^4$  to  $2 \times 10^5$  cells in 5  $\mu$ l of PBS, and the *c-kit*<sup>NEG</sup> cells were suspended at a concentration of  $5 \times 10^4$  to  $5 \times 10^5$  cells in 5  $\mu$ l of PBS<sup>10</sup>.

### Myocardial infarction

Myocardial infarction was induced in female C57BL/6 mice at 2 months of age as described<sup>11</sup>; 3–5 h after infarction, the thorax was re-opened and 2.5  $\mu$ l PBS containing Lin<sup>-</sup>*c-kit*<sup>POS</sup> cells were injected in the anterior and posterior aspects of the viable myocardium bordering the infarct. Infarcted mice that were not injected or injected with Lin<sup>-</sup>*c-kit*<sup>NEG</sup> cells and sham-operated mice were used as controls. All animals were killed 9  $\pm$  2 days after surgery. Protocols were approved by an institutional review board.

### Ventricular function

Mice were anaesthetized with chloral hydrate (400 mg per kg (body weight)), intraperitoneally (i.p.), and the right carotid artery was cannulated with a microtip pressure transducer (model SPR-671; Millar) for the measurements of LV pressures, and LV + and LV - dP/dt in the closed-chest preparation. The abdominal aorta was cannulated, the heart was arrested in diastole, and the myocardium was perfused retrogradely with 10% buffered formalin<sup>11,26</sup>. Three tissue sections, from the base to the apex of the left ventricle, were stained with haematoxylin and eosin. At 9  $\pm$  2 days after coronary occlusion, the infarcted portion of the ventricle was easily identifiable grossly and histologically (see Fig. 1a). The lengths of the endocardial and epicardial surfaces delimiting the infarcted region, and the endocardium and epicardium of the entire left ventricle, were measured in each section. Subsequently, their quotients were computed to yield the average infarct size in each case. This was accomplished at  $\times 4$  magnification with an image analyser connected to a microscope<sup>11</sup>.

# Cell proliferation and EGFP detection

To establish whether Lin<sup>-</sup>c-kit<sup>POS</sup> cells resulted in myocardial regeneration, we administered BrdU (50 mg per kg (body weight), i.p.) to the animals daily for 4–5 consecutive days before death. Sections were incubated with anti-BrdU antibody, and BrdU labelling of cardiac cells was measured<sup>17</sup>. Moreover, expression of Ki67 in nuclei was evaluated by treating samples with a rabbit polyclonal anti-mouse Ki67 antibody (Dako). Fluorescein isothiocyanate (FITC)-conjugated goat anti-rabbit IgG was used as secondary antibody. EGFP was detected with a rabbit polyclonal anti-GFP (Molecular Probes). Myocytes were recognized with a mouse monoclonal anti-cardiac myosin heavy chain (MAB 1552; Chemicon) or a mouse monoclonal anti sarcomeric  $\alpha$ -actin (clone 5C5; Sigma), endothelial cells with a rabbit polyclonal anti-human factor VIII (Sigma) and smooth muscle cells with a mouse monoclonal anti-smooth-muscle  $\alpha$ -actin (clone 1A4; Sigma). Nuclei were stained with propidium iodide, 10  $\mu$ g ml<sup>-1</sup> (refs 27, 28). We determined the percentages of myocyte (M), endothelial cell (EC) and smooth muscle cell (SMC) nuclei labelled by BrdU and Ki67 by confocal microscopy. This was accomplished by dividing the number of nuclei labelled by the total number of nuclei examined. Numbers of nuclei sampled in each cell population were as follows. BrdU labelling: M, 2,908; EC, 2,153; SMC, 4,877. Ki67 labelling: M, 3,771; EC, 4,051; SMC, 4,752. Numbers of cells counted for EGFP labelling: M, 3,278; EC, 2,056; SMC, 1,274. We determined the percentage of myocytes in the regenerating myocardium by delineating the area occupied by cardiac-myosin-stained cells and dividing this by the total area represented by the infarcted region in each case.

# Y chromosome

For the fluorescence *in situ* hybridization assay, we exposed sections to a denaturing solution containing 70% formamide. After dehydration with ethanol, sections were hybridized with the DNA probe CEP Y (satellite III) Spectrum Green (Vysis) for 3 h (ref. 29). Nuclei were stained with propidium iodide.

# Transcription factors and connexin 43

Sections were incubated with rabbit polyclonal anti-MEF2 (C-21; Santa Cruz), rabbit polyclonal anti-GATA-4 (H-112; Santa Cruz), rabbit polyclonal anti-Csx/Nkx2.5 (obtained from Dr S. Izumo) and rabbit polyclonal anti-connexin 43 (Sigma). We used FITC-conjugated goat anti-rabbit IgG (Sigma) as the secondary antibody<sup>30</sup>.

# Statistical analysis

Results are presented as means  $\pm$  s.d. Significance between two measurements was determined by Student's *t*-test, and in multiple comparisons was evaluated by the Bonferroni method. Values of *P* < 0.05 were considered significant.

Received 7 November 2000; accepted 16 February 2001.

1. Pfeffer, M. A. & Braunwald, E. Ventricular remodeling after myocardial infarction. *Circulation* **81**, 1161–1172 (1990).
2. Ferrari, G. *et al.* Muscle regeneration by bone marrow-derived myogenic progenitors. *Science* **279**, 1528–1530 (1998).
3. Jackson, K. A., Mi, T. & Goodell, M. A. Hematopoietic potential of stem cells isolated from murine skeletal muscle. *Proc. Natl Acad. Sci. USA* **96**, 14482–14486 (1999).
4. Eglitis, M. A. & Mezey, E. Hematopoietic cells differentiate into both microglia and macroglia in the brains of adult mice. *Proc. Natl Acad. Sci. USA* **94**, 4080–4085 (1997).
5. Theise, N. D. *et al.* Liver from bone marrow in humans. *Hepatology* **32**, 11–16 (2000).
6. Brazelton, T. A., Rossi, F. M. V., Keshet, G. I. & Blau, H. M. From marrow to brain: expression of neuronal phenotypes in adult mice. *Science* **290**, 1775–1779 (2000).
7. Mezey, E., Chandross, K. J., Harta, G., Maki, R. A. & McKeacher, S. R. Turning blood into brain: cells bearing neuronal antigens generated in vivo from bone marrow. *Science* **290**, 1779–1782 (2000).
8. Vogel, G. Stem cells: new excitations, persistent questions. *Science* **290**, 1672–1674 (2000).
9. Hadjantonakis, A. K., Gerstenstein, M., Ikawa, M., Okabe, M. & Nagy, A. Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mech. Dev.* **76**, 79–90 (1998).
10. Orlie, D., Fischer, R., Nishikawa, S.-I., Nienhuis, A. W. & Bodine, D. M. Purification and characterization of heterogeneous pluripotent hematopoietic stem cell populations expressing high levels of *c-kit* receptor. *Blood* **82**, 762–770 (1993).
11. Li, Q. *et al.* Overexpression of insulin-like growth factor-1 in mice protects from myocyte death after infarction, attenuating ventricular dilation, wall stress, and cardiac hypertrophy. *J. Clin. Invest.* **100**, 1991–1999 (1997).
12. Durocher, D., Charron, F., Warren, R., Schwartz, R. J. & Nemer, M. The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J.* **16**, 5687–5696 (1997).
13. Morin, S., Charron, F., Robitaille, L. & Nemer, M. GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J.* **19**, 2046–2055 (2000).
14. Beardslee, M. A., Laing, J. G., Beyer, E. C. & Saffitz, J. E. Rapid turnover of connexin43 in the adult rat heart. *Circ. Res.* **83**, 629–635 (1998).
15. Musil, L. S., Le, A. N., VanSlyke, J. K. & Roberts, L. M. Regulation of connexin degradation as a mechanism to increase gap junction assembly and function. *J. Biol. Chem.* **275**, 25207–25215 (2000).
16. Scholzen, T. & Gerdes, J. The ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* **182**, 311–322 (2000).
17. Kajstura, J. *et al.* Myocyte cellular hyperplasia and myocyte cellular hypertrophy contribute to chronic ventricular remodeling in coronary artery narrowing-induced cardiomyopathy in rats. *Circ. Res.* **74**, 383–400 (1994).
18. Clarke, D. L. Generalized potential of adult neural stem cells. *Science* **288**, 1660–1663 (2000).

19. Theise, N. D. *et al.* Derivation of hepatocytes from bone marrow cells in mice after radiation-induced myeloablation. *Hepatology* **31**, 235–240 (2000).
20. Lagasse, E. *et al.* Purified hematopoietic stem cells can differentiate into hepatocytes *in vivo*. *Nature Med.* **6**, 1229–1234 (2000).
21. Kunisada, T. *et al.* Transgene expression of steel factor in the basal layer of epidermis promotes survival, proliferation, differentiation and migration of melanocyte precursors. *Development* **125**, 2915–2923 (1998).
22. Matsui, Y., Zsebo, K. M. & Hogan, B. Embryonic expression of a hematopoietic growth factor encoded by the S1 locus and the ligand for c-kit. *Nature* **347**, 667–669 (1990).
23. Morrison, S. J., Uchida, N. & Weissman, I. L. The biology of hematopoietic stem cells. *Annu. Rev. Cell Dev. Biol.* **11**, 35–71 (1994).
24. Morrison, S. J., Shah, N. M. & Anderson, D. J. Regulatory mechanisms in stem cell biology. *Cell* **88**, 287–298 (1997).
25. Tomita, S. *et al.* Autologous transplantation of bone marrow cells improves damaged heart function. *Circulation* **100** (Suppl.), II-247–II-256 (1999).
26. Li, B. *et al.* Insulin-like growth factor-1 attenuates the detrimental impact of non occlusive coronary artery constriction on the heart. *Circ. Res.* **84**, 1007–1019 (1999).
27. Leri, A. *et al.* Stretch-mediated release of angiotensin II induces myocyte apoptosis by activating p53 that enhances the local renin-angiotensin system and decreases the Bcl-2-to-Bax protein ratio in the cell. *J. Clin. Invest.* **101**, 1326–1342 (1998).
28. Kajstura, J. *et al.* Myocyte proliferation in end-stage cardiac failure in humans. *Proc. Natl Acad. Sci. USA* **95**, 8801–8805 (1998).
29. Kajstura, J. *et al.* Telomere shortening is an *in vivo* marker of myocyte replication and aging. *Am. J. Pathol.* **156**, 813–819 (2000).
30. Leri, A. *et al.* Pacing-induced heart failure in dogs enhances the expression of p53 and p53-dependent genes in ventricular myocytes. *Circulation* **97**, 194–203 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

# Acknowledgements

We thank Dr S. Izumo for providing us with the Csx2.5 antibody. This work was supported by grants from the NIH. S.C. is supported by a fellowship from the Mario Negri Institute of Pharmacologic Research, Milan, Italy.

Correspondence and requests for materials should be addressed to P.A. (e-mail: [piero\\_anversa@nymc.edu](mailto:piero_anversa@nymc.edu)).

# CaT1 manifests the pore properties of the calcium-release-activated calcium channel

Lixia Yue<sup>‡</sup>, Ji-Bin Peng<sup>†</sup>, Matthias A. Hediger<sup>†</sup> & David E. Clapham<sup>‡</sup>

<sup>‡</sup> Howard Hughes Medical Institute and \* Children's Hospital, Harvard Medical School, Enders 1309, 320 Longwood Avenue, Boston, Massachusetts 02115, USA  
<sup>†</sup> Membrane Biology Program and Renal Division, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

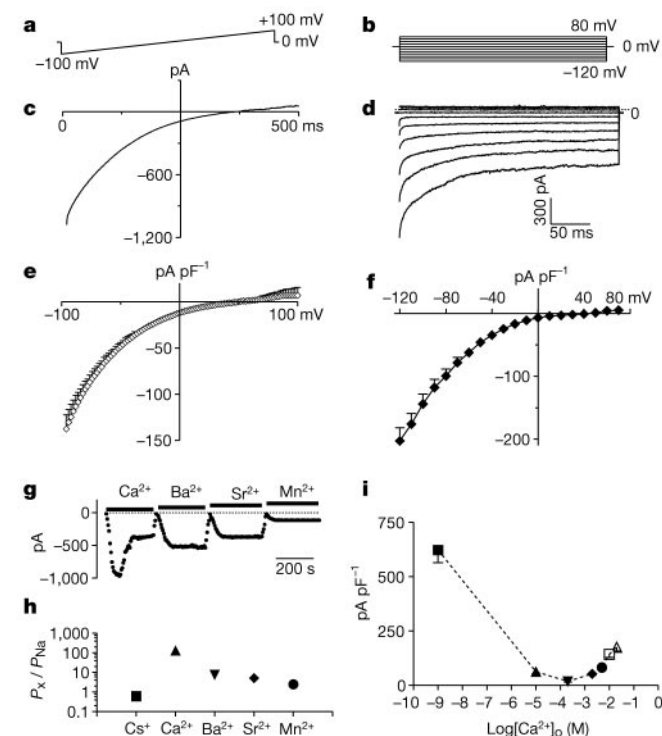
The calcium-release-activated Ca<sup>2+</sup> channel, I<sub>CRAC</sub><sup>1–3</sup>, is a highly Ca<sup>2+</sup>-selective ion channel that is activated on depletion of either intracellular Ca<sup>2+</sup> levels or intracellular Ca<sup>2+</sup> stores. The unique gating of I<sub>CRAC</sub> has made it a favourite target of investigation for new signal transduction mechanisms; however, without molecular identification of the channel protein, such studies have been inconclusive. Here we show that the protein CaT1 (ref. 4), which has six membrane-spanning domains, exhibits the unique biophysical properties of I<sub>CRAC</sub> when expressed in mammalian cells. Like I<sub>CRAC</sub>, expressed CaT1 protein is Ca<sup>2+</sup> selective, activated by a reduction in intracellular Ca<sup>2+</sup> concentration, and inactivated by higher intracellular concentrations of Ca<sup>2+</sup>. The channel is indistinguishable from I<sub>CRAC</sub> in the following features: sequence of selectivity to divalent cations; an anomalous mole fraction effect; whole-cell current kinetics; block by lanthanum; loss of selectivity in the absence of divalent cations; and single-channel conductance to Na<sup>+</sup> in divalent-ion-free conditions. CaT1 is activated by both passive and active depletion of calcium stores. We propose that CaT1 comprises all or part of the I<sub>CRAC</sub> pore.



Calcium-selective ion channels markedly increase cytosolic concentrations of  $\text{Ca}^{2+}$  ( $[\text{Ca}^{2+}]_i$ ), the trigger for cellular events ranging from contraction to secretion. The highly selective  $\text{Ca}^{2+}$  channels include the voltage-dependent channels and  $I_{\text{CRAC}}^{1-3}$ . The role of  $I_{\text{CRAC}}$  is best defined in unexcitable cells, where hyperpolarization increasingly draws  $\text{Ca}^{2+}$  into the cell. Commonly, a store-operated  $\text{Ca}^{2+}$  entry process<sup>5</sup> is initiated by a large receptor-mediated transient  $\text{Ca}^{2+}$  release from the endoplasmic reticulum, followed by a prolonged plateau dependent on extracellular  $[\text{Ca}^{2+}]$  ( $[\text{Ca}^{2+}]_o$ ) through  $I_{\text{CRAC}}^{6,7}$ . Identifying the  $I_{\text{CRAC}}$  channel protein will speed the understanding of two significant mysteries: the identity of the signal from the depleted stores that gates  $\text{Ca}^{2+}$  entry; and the biological role of the nearly ubiquitous  $I_{\text{CRAC}}$  current.

In CaT1-transfected Chinese hamster ovary (CHO)-K1 cells, a large inwardly rectifying current was elicited by a linear voltage ramp when intracellular  $[\text{Ca}^{2+}]$  ( $[\text{Ca}^{2+}]_i$ ) was buffered by 10 mM EGTA in the patch pipette. Mock-transfected cells exhibited only a small background current. The CaT1 current reversed at  $+49 \pm 5$  mV (mean  $\pm$  s.e.m.;  $n = 12$ ; Fig. 1). Currents evoked by voltage steps in CaT1-expressing cells showed substantial time-dependent inactivation at potentials negative to  $-40$  mV. The inactivation process at  $-100$  mV was best fit by two exponential functions with time constants  $\tau_1$  and  $\tau_2$  ( $\tau_1 = 10.8 \pm 1.4$  ms and  $\tau_2 = 58.2 \pm 7.2$  ms;  $n = 6$ ). The  $I-V$  relations obtained from both ramp and voltage steps rectified sharply in the inward direction, similar to those reported previously for  $I_{\text{CRAC}}$  current<sup>2,7-9</sup>.

The relative conductances of CaT1 were  $\text{Ca}^{2+} > \text{Ba}^{2+} > \text{Sr}^{2+} > \text{Mn}^{2+}$ , which match those shown for  $I_{\text{CRAC}}^{7,10,11}$  (Fig. 1g). CaT1

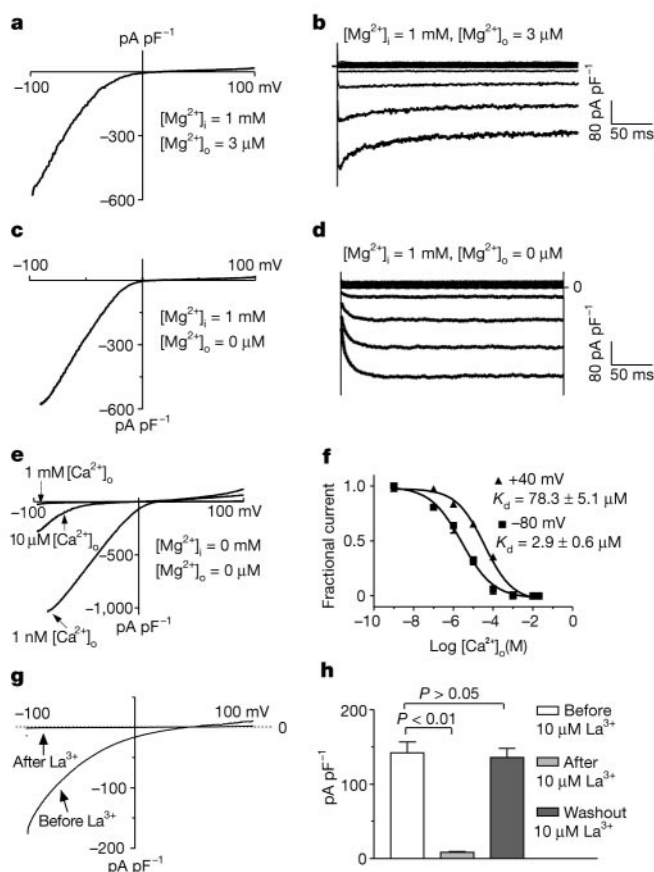


**Figure 1** Currents in CaT1-expressing CHO-K1 cells. **a**, The 500-ms voltage ramp ( $-100$  to  $+100$  mV) for **c** and **e**.  $V_{\text{h}} = 0$  mV. **b**, Voltage protocol for **d** and **f**, with 1 s between steps. **c**, Ramp current (reversal,  $+50$  mV). **d**, Step current. **e**, Current densities from CaT1-transfected cells (mean  $\pm$  s.e.m.;  $n = 12$ ). **f**,  $I-V$  relation (mean  $\pm$  s.e.m.;  $n = 12$ ). **g**, Relative conductances of CaT1 to 10 mM  $\text{Ca}^{2+}$ ,  $\text{Ba}^{2+}$ ,  $\text{Sr}^{2+}$  and  $\text{Mn}^{2+}$  (10  $\mu\text{M}$   $[\text{Ca}^{2+}]$  between solution changes). **h**, Relative permeability (to  $\text{Na}^+$ ) of  $\text{Cs}^+$  (0.6),  $\text{Ca}^{2+}$  (130),  $\text{Ba}^{2+}$  (7.5),  $\text{Sr}^{2+}$  (5.1) and  $\text{Mn}^{2+}$  (2.4) through CaT1 channels ( $n = 8$  for each). **i**, CaT1 anomalous mole-fraction behaviour (mean  $\pm$  s.e.m.;  $n = 6$ ) at  $-100$  mV.  $[\text{Ca}^{2+}]$  in DVF solution was less than 1 nM.

current rapidly inactivates in  $[\text{Ca}^{2+}]_o$ , but not substantially in  $\text{Ba}^{2+}$  or  $\text{Sr}^{2+}$ —a well described property of the  $I_{\text{CRAC}}$  current<sup>10</sup>. To study the relative divalent permeabilities of CaT1, we substituted equimolar  $\text{Ba}^{2+}$ ,  $\text{Sr}^{2+}$  or  $\text{Mn}^{2+}$  for 10 mM  $\text{Ca}^{2+}$  in the extracellular bathing solution, and measured reversal potentials. CaT1 current amplitudes were substantially decreased after substituting  $\text{Ca}^{2+}$  with other divalent cations. CaT1 is most permeable to  $\text{Ca}^{2+}$  and least permeable to  $\text{Mn}^{2+}$  in the order  $\text{Ca}^{2+} \gg \text{Ba}^{2+} > \text{Sr}^{2+} > \text{Mn}^{2+}$  with a permeability ( $P$ ) ratio,  $P_{\text{Ca}}/P_{\text{Na}} \approx 130$  (Fig. 1h). This relative permeability sequence is unlike that of the voltage-dependent channels  $\text{Ca}_v$  1.1–1.3, in which  $P_{\text{Ba}} \geq P_{\text{Ca}}$  (refs 2, 12).

Calcium-selective channels exhibit a concentration-dependent permeability ratio, which is thought to be a consequence of the  $\text{Ca}^{2+}$  channel's capacity to hold two or more divalent ions in the pore at the same time<sup>12</sup>. The  $[\text{Ca}^{2+}]_o$  dependence of CaT1 current densities illustrates the anomalous mole fraction behaviour observed for highly  $\text{Ca}^{2+}$ -selective channels (Fig. 1i).

One of the hallmarks of  $I_{\text{CRAC}}$  is its increased permeability to



**Figure 2** CaT1 monovalent currents. **a**, CaT1 currents elicited by voltage ramps in 3  $\mu\text{M}$   $[\text{Mg}^{2+}]_o$  DVF solution. **b**, Monovalent  $\text{Na}^+$  currents through CaT1 induced by voltage steps in 3  $\mu\text{M}$   $[\text{Mg}^{2+}]_o$  DVF solution. **c**, CaT1 currents elicited by voltage ramp in 0  $\mu\text{M}$   $[\text{Mg}^{2+}]_o$  DVF solution. **d**, CaT1 currents induced by voltage steps in 0  $\mu\text{M}$   $[\text{Mg}^{2+}]_o$  DVF solution. **e**, Inhibition of monovalent CaT1 currents by 10  $\mu\text{M}$  and 1 mM  $[\text{Ca}^{2+}]_o$  ( $\text{Mg}^{2+}$ -free pipette solution, ramp protocol). **f**, Dose-response curves of  $\text{Ca}^{2+}$  blockade on monovalent currents at  $-80$  and  $+40$  mV. Current amplitudes at  $-80$  and  $+40$  mV were normalized to values at 1 nM  $\text{Ca}^{2+}$ , and fitted according to  $y = 1/(1 + x/K_d)^n$ , where  $x$  is the external  $[\text{Ca}^{2+}]_o$ ,  $y$  is the fractional current,  $K_d$  is the apparent dissociation constant and  $n$  is the Hill coefficient.  $K_d$  was  $2.9 \pm 0.6 \mu\text{M}$  at  $-80$  mV, and  $78.3 \pm 5.1 \mu\text{M}$  at  $+40$  mV (mean  $\pm$  s.e.m.; both  $n = 5$ ). **g**, CaT1 currents recorded with the ramp protocol before and after 10  $\mu\text{M}$   $\text{La}^{3+}$  in 10 mM  $[\text{Ca}^{2+}]_o$ . **h**, Current densities (mean  $\pm$  s.e.m.) measured at  $-100$  mV in 10 mM  $[\text{Ca}^{2+}]_o$  before and after 10  $\mu\text{M}$   $\text{La}^{3+}$  addition to the bath ( $n = 6$ ).

monovalent ions, such as  $\text{Na}^+$ , when divalents are completely removed from the bathing solution. The conversion to non-selectivity is thought to result from the removal of divalents from binding sites (glutamic or aspartic acid residues) lining the calcium-selective pore. When CaT1-expressing cells were exposed to divalent-free (DVF) solution with free  $\text{Mg}^{2+}$  buffered to  $3 \mu\text{M}$  (ref. 8), currents induced by voltage ramps (Fig. 2a) and voltage steps (Fig. 2b) were both increased markedly. The average current density at  $-100 \text{ mV}$  in DVF solution was increased about fivefold compared with that in  $10 \text{ mM Ca}^{2+}$  solution, indicating that CaT1-encoded channels are permeable to  $\text{Na}^+$  in the presence of lowered external divalent cations.

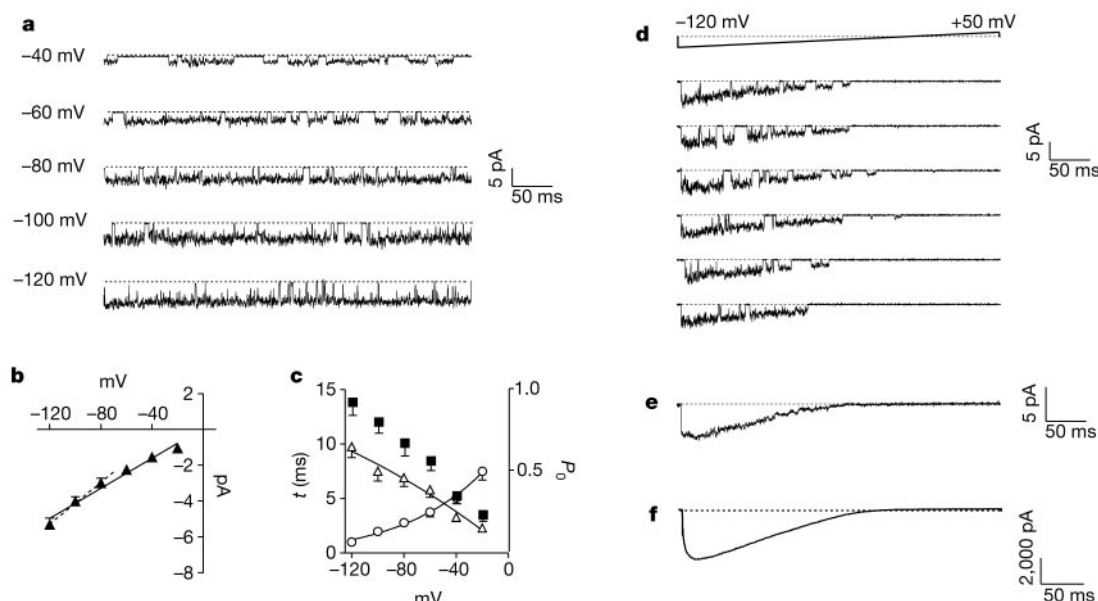
Notably, monovalent CaT1 current amplitude showed time-dependent changes in DVF with  $0 \mu\text{M}$  free  $[\text{Mg}^{2+}]_o$  (Fig. 2d). This slower development ( $\tau_{on} = 6.3 \pm 0.6 \text{ ms}$  at  $-100 \text{ mV}$ ;  $n = 6$ ) of the CaT1 current in the near complete absence of divalent ions had not been tested before for  $I_{\text{CRAC}}$ , but indicates that divalent ions occupying the pore of the channel alter the voltage-dependent activation process. (In preliminary experiments in rat basophilic leukaemia (RBL) cells, a delay is observed when  $I_{\text{CRAC}}$  is measured under DVF conditions.) The monovalent current amplitude of CaT1 currents was 10-fold larger than currents in  $10 \text{ mM } [\text{Ca}^{2+}]_o$  when  $\text{Mg}^{2+}$  was removed from the pipette solution (Fig. 2e). Monovalent currents recorded with a  $\text{Mg}^{2+}$ -free pipette were blocked by 50% with  $2.9 \pm 0.6 \mu\text{M } [\text{Ca}^{2+}]_o$  at  $-80 \text{ mV}$  and  $78.3 \mu\text{M } [\text{Ca}^{2+}]_o$  at  $+40 \text{ mV}$  (Fig. 2f), suggesting a voltage-dependent block mechanism and consistent with the behaviour of  $I_{\text{CRAC}}$ <sup>9</sup>. CaT1 currents recorded in  $10 \text{ mM Ca}^{2+}$  were completely and reversibly blocked by  $10 \mu\text{M La}^{3+}$  (Fig. 2g, h), as reported for  $I_{\text{CRAC}}$ <sup>13–15</sup>.

The single-channel conductance of  $I_{\text{CRAC}}$  has been reported to increase from an estimated  $0.5 \text{ pS}$  to  $40 \text{ pS}$  in the absence of divalent ions<sup>8</sup>. We recorded CaT1 single-channel currents under similar conditions to those described<sup>8</sup> (Fig. 3). No similar single-channel openings were observed in mock-transfected cells under the same experimental conditions. A linear regression fit of the current amplitude as a function of voltage yielded a slope conductance of  $42 \pm 5 \text{ pS}$  ( $n = 6$ ) between  $-120$  and  $-20 \text{ mV}$ , and  $58 \pm 4 \text{ pS}$  ( $n = 6$ )

between  $-120$  and  $-80 \text{ mV}$ . The latter is close to the slope conductance reported for  $I_{\text{CRAC}}$  ( $55 \text{ pS}$  between  $-120$  and  $-80 \text{ mV}$ )<sup>16</sup>. Similar CaT1 single-channel conductances were obtained in outside-out patches (data not shown). CaT1 exhibited a high open probability ( $P_o$ ) at negative potentials ( $0.9 \pm 0.1$  at  $-120 \text{ mV}$ ), and  $P_o$  decreased at more depolarized potentials ( $0.7 \pm 0.1$  at  $-80 \text{ mV}$ ; Fig. 3c). The high  $P_o$  was correlated with longer open times and brief closed times at negative potentials. For example,  $\tau_o$  was  $9.8 \text{ ms}$  and  $\tau_c$  was  $1.0 \text{ ms}$  at  $-120 \text{ mV}$ , whereas  $\tau_o$  decreased to  $6.3 \text{ ms}$  and  $\tau_c$  increased to  $2.8 \text{ ms}$  at  $-80 \text{ mV}$  (Fig. 3c). Similarly, the single-channel open probability for  $I_{\text{CRAC}}$  was  $0.95$  at  $-120 \text{ mV}$  and decreased to  $0.8$  at  $-80 \text{ mV}$  (ref. 16). (At more depolarized potentials, CaT1  $P_o$  differs from that reported for  $I_{\text{CRAC}}$ ; however,  $P_o$  varies with the cell and mode of gating, and similar  $P_o$  values (such as  $0.55$  at  $-40 \text{ mV}$ ) have been recorded; Fomina, A. F., Kerschbaum, H. H. and Cahalan, M. D.; personal communication). The ensemble of single-channel currents averaged from 40 ramp protocol sweeps (Fig. 3d, e) exhibited inward rectification that was indistinguishable from that of the whole-cell current (Fig. 3f).

We examined the store-depletion properties of CaT1 expressed in CHO-K1 cells. We used CHO-K1 cells in these expression studies because native  $I_{\text{CRAC}}$  currents are not detectable ( $< 0.1 \text{ pA pF}^{-1}$ ). As observed in many recordings of  $I_{\text{CRAC}}$ , when cells were equilibrated in  $10 \text{ mM } [\text{Ca}^{2+}]_o$  passive store depletion by strong buffering of  $[\text{Ca}^{2+}]_i$  ( $10 \text{ mM EGTA}$ ) evoked large CaT1 currents (Fig. 4a), whereas weak buffering ( $0.05 \text{ mM EGTA}$ ) failed to activate CaT1 (Fig. 4b). In  $10 \text{ mM } [\text{Ca}^{2+}]_o$ , the CaT1 currents reached a peak in about  $30 \text{ s}$  and inactivated slowly (Fig. 4a, g).  $I_{\text{CRAC}}$  displays a similar slow inactivation process caused by store refilling<sup>10,17</sup> or negative feedback owing to  $\text{Ca}^{2+}$  entry<sup>17</sup>. Also, like  $I_{\text{CRAC}}$ <sup>17</sup>, the slow inactivation was reversed on removing extracellular  $\text{Ca}^{2+}$ . The current completely recovered after a 2-min exposure to  $\text{Ca}^{2+}$ -free ( $\sim 10 \mu\text{M } [\text{Ca}^{2+}]_o$ ) solution (Fig. 4a).

Previous studies have shown that lowering  $[\text{Ca}^{2+}]_o$  depletes stores in many types of cells<sup>18,19</sup>. We found that incubation of cells in nominally free  $\text{Ca}^{2+}$  ( $10 \mu\text{M Ca}^{2+}$ ; Fig. 4c, d) or DVF ( $\sim 1 \text{ nM Ca}^{2+}$ ) solution (Fig. 4e, f) activated CaT1 in cells with either strongly or weakly buffered  $[\text{Ca}^{2+}]_i$ . Three pieces of evidence show



**Figure 3** CaT1 single-channel properties (inside-out patch). **a**, Single-channel currents evoked at the indicated voltages ( $V_H = 0 \text{ mV}$ ; DVF solution). **b**, Unitary  $I$ - $V$  relation of CaT1 (mean  $\pm$  s.e.m.;  $n = 6$ ). Slope conductance was  $42 \text{ pS}$  from  $-120$  to  $-20 \text{ mV}$  (solid line); and  $58 \text{ pS}$  from  $-120$  to  $-80 \text{ mV}$  (dashed lines). **c**, Open probability ( $P_o$ , squares), and

open ( $\tau_o$ , triangles) and closed ( $\tau_c$ , circles) time constants of CaT1 single channels plotted against voltage (mean  $\pm$  s.e.m.;  $n = 6$ ). **d**, CaT1 single-channel currents elicited by 500-ms ramps in  $140 \text{ mM K}^+$ , DVF solution. **e**, Ensemble-averaged current (40 traces). **f**, Macroscopic current recorded with the same voltage protocol as in **e**.

that passive store depletion activated monovalent CaT1 current (Fig. 4c–f).

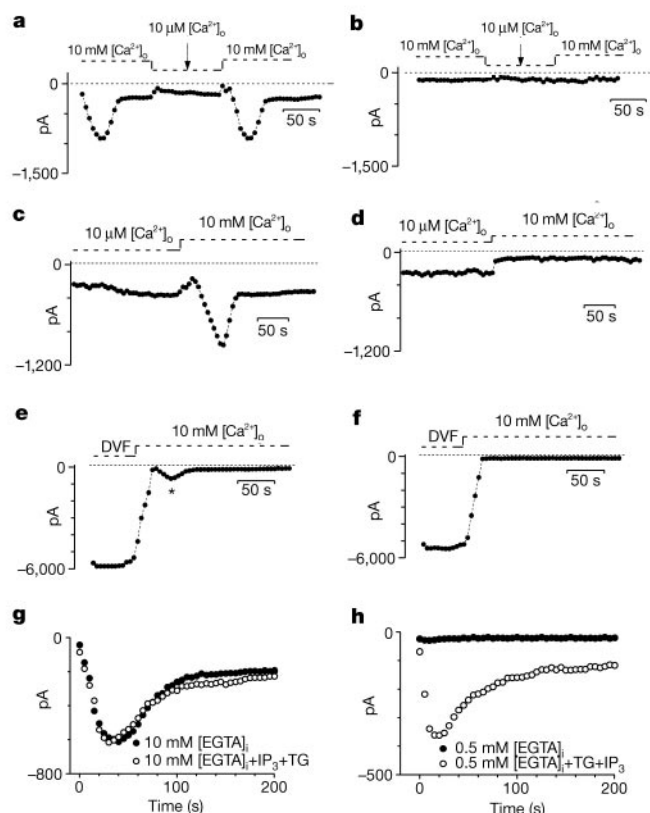
First, CaT1 current amplitude reached maximal levels immediately after the onset of whole-cell recording (Fig. 4c–f) when stores had been passively depleted, whereas current slowly developed over 30–50 s if stores were not depleted before whole-cell perfusion (Fig. 4a). Second, maximal monovalent CaT1 current amplitudes were the same in passively depleted cells whether weak or strong intracellular buffers were used ( $-63 \pm 8 \text{ pA pF}^{-1}$  (Fig. 4c) versus  $-60 \pm 7 \text{ pA pF}^{-1}$  (Fig. 4d); and  $-660 \pm 7 \text{ pA pF}^{-1}$  (Fig. 4e) versus  $-610 \pm 58 \text{ pA pF}^{-1}$  (Fig. 4f);  $n = 6$  for each group;  $P > 0.05$  for each comparison). Finally,  $\text{Ca}^{2+}$  currents were induced when  $[\text{Ca}^{2+}]_o$  was increased from  $10 \mu\text{M}$  (nominally  $\text{Ca}^{2+}$  free) or  $1 \text{ nM}$  DVF to  $10 \text{ mM}$   $\text{Ca}^{2+}$  in cells in which  $[\text{Ca}]_i$  was strongly buffered ( $10 \text{ mM}$  EGTA; Fig. 4c, e) but not in weakly buffered cells ( $0.05 \text{ mM}$  EGTA;

Fig. 4d, f). Presumably  $[\text{Ca}^{2+}]_i$  stores are not depleted in cells in which  $[\text{Ca}^{2+}]_i$  is weakly buffered.

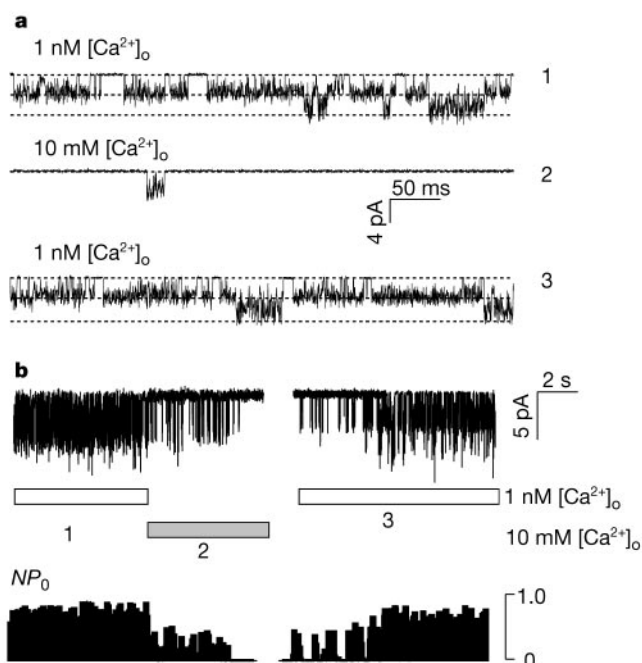
As for the whole-cell currents, CaT1 single-channel currents in cell-attached patches were activated by equilibration in DVF solution (Fig. 5a), but not in cells continuously perfused with  $10 \text{ mM}$   $\text{Ca}^{2+}$  (data not shown). Exposure to  $10 \text{ mM}$   $[\text{Ca}^{2+}]_o$  decreased channel activity and eventually completely inactivated single channels in cell-attached patches. The inactivation induced by  $[\text{Ca}^{2+}]_o$  was reversed on equilibrating cells in DVF (Fig. 5a, b). These results show that passive store depletion or lowering  $[\text{Ca}^{2+}]_i$  activated CaT1.

Active intracellular  $\text{Ca}^{2+}$  store depletion is experimentally implemented by intracellular application of inositol-1,4,5-triphosphate ( $\text{InsP}_3$ ) to release  $\text{Ca}^{2+}$  from stores directly, or application of thapsigargin to block smooth endoplasmic reticular  $\text{Ca}^{2+}$ -ATPase pumps. CaT1 was reportedly not activated by active store depletion<sup>4</sup>, and indeed its sensitivity to active store depletion is not at first obvious in cells overexpressing CaT1. Because  $[\text{Ca}^{2+}]_i$  levels affect store depletion and  $I_{\text{CRAC}}$  activation<sup>20</sup>, we optimized our experimental conditions by comparing CaT1 currents in which  $0.5 \text{ mM}$  EGTA or  $10 \text{ mM}$  EGTA was dialysed in cells bathed in  $10 \text{ mM}$   $[\text{Ca}^{2+}]_o$ .

A principal assumption in our experiments is that CaT1 does not comprise the signal transduction mechanism that senses store depletion itself, and thus CaT1 overexpression in CHO-K1 cells may not be matched by complete upregulation of the signal transduction apparatus. Thus, to try to limit this mismatch, we recorded currents within 8–12 h after transfection, when CaT1 expression levels were presumably lower. Under these conditions, peak CaT1 currents activated by  $10 \text{ mM}$  EGTA averaged  $-89 \pm 8 \text{ pA pF}^{-1}$  (Fig. 4g,  $n = 8$ ). As stores were presumably fully depleted under these conditions, inclusion of  $\text{InsP}_3$  and thapsigargin did not induce additional current beyond that activated



**Figure 4** Activation of CaT1 currents. **a, b**, Time-dependent development and inactivation of CaT1 currents in cells dialysed with  $10 \text{ mM}$  EGTA (**a**) or  $0.05 \text{ mM}$  EGTA (**b**), when continuously superfused with  $10 \text{ mM}$   $[\text{Ca}^{2+}]_o$ . Peak current amplitude at  $-100 \text{ mV}$  was plotted. Current amplitude (**a**) was maximal within 30 s after break-in and gradually inactivated. Inactivation of CaT1 was reversible on exposure to  $\text{Ca}^{2+}$ -free solution. No currents were activated in cells dialysed with  $0.05 \text{ mM}$  EGTA. **c, d**, Time-dependent changes of CaT1 currents recorded in cells dialysed with  $10 \text{ mM}$  EGTA (**c**) or  $0.05 \text{ mM}$  EGTA (**d**). Cells were equilibrated in  $10 \mu\text{M}$   $\text{Ca}^{2+}$ . Small monovalent currents were observed after establishing the whole-cell configuration. After superfusion with  $10 \text{ mM}$   $\text{Ca}^{2+}$  solution, large  $\text{Ca}^{2+}$  currents with similar activation and inactivation properties were induced in cells dialysed with  $10 \text{ mM}$  EGTA (**c**), but not  $0.05 \text{ mM}$  EGTA (**d**). **e, f**, Time-dependent changes of CaT1 currents in cells in DVF solution. Monovalent currents were elicited in cells dialysed with  $10 \text{ mM}$  EGTA (**e**) or  $0.05 \text{ mM}$  EGTA (**f**). After superfusion with  $10 \text{ mM}$   $\text{Ca}^{2+}$  solution, only the cells dialysed with  $10 \text{ mM}$  EGTA displayed the  $\text{Ca}^{2+}$  current (asterisk). **g**, CaT1 currents induced by 100-ms voltage ramps in cells dialysed with  $10 \text{ mM}$  EGTA or  $10 \text{ mM}$  EGTA +  $20 \mu\text{M}$   $\text{InsP}_3$  +  $2 \mu\text{M}$  thapsigargin in cells 8 h after transfection. Peak current amplitude at  $-100 \text{ mV}$  is plotted. **h**, CaT1 currents recorded in cells dialysed with  $0.5 \text{ mM}$  EGTA or  $0.5 \text{ mM}$  EGTA +  $20 \mu\text{M}$   $\text{InsP}_3$  +  $2 \mu\text{M}$  thapsigargin elicited by 100-ms voltage ramps.



**Figure 5** Single-channel CaT1 currents. **a**, Blockade of single-channel currents by  $[\text{Ca}^{2+}]_o$ . Single-channel currents recorded in a cell-attached patch at  $-100 \text{ mV}$  in DVF ( $1 \text{ nM}$   $\text{Ca}^{2+}$ ; 1), after  $10 \text{ mM}$   $\text{Ca}^{2+}$ , and after return to  $1 \text{ nM}$   $\text{Ca}^{2+}$ . After superfusion of  $10 \text{ mM}$   $\text{Ca}^{2+}$ , channel activity declined markedly (2) and reversibly (3). **b**, Continuously recorded CaT1 single-channel currents at  $-100 \text{ mV}$  in DVF ( $1 \text{ nM}$   $\text{Ca}^{2+}$ ), after  $10 \text{ mM}$   $\text{Ca}^{2+}$ , and after return to  $1 \text{ nM}$   $\text{Ca}^{2+}$ .  $N$ , number of channels;  $P_o$ , probability of opening.



with 10 mM EGTA alone (Fig. 4g), and displayed a similar time course of activation to peak CaT1 current (between 10 and 30 s).

With low intracellular  $\text{Ca}^{2+}$  buffer (0.5 mM EGTA dialysed cells), CaT1 was not activated (Fig. 4h, top trace). When  $\text{InsP}_3$  and thapsigargin were included under these conditions, however, CaT1 was activated (mean current density of  $-49 \pm 7 \text{ pA pF}^{-1}$ ;  $n = 6$ ; Fig. 4h, bottom trace). Furthermore, when  $[\text{Ca}^{2+}]_i$  measurements were performed with ratiometric Fura-2 imaging, CaT1-mediated entry was stimulated by  $2 \mu\text{M}$  thapsigargin alone (data not shown). The levels of CaT1 current are roughly 400–500-fold above a putative, intrinsic, maximally activated  $I_{\text{CRAC}}$  current ( $0.1 \text{ pA pF}^{-1}$ ), and support our conclusion that CaT1 is activated by store depletion but is limited in extent by the native signal transduction apparatus.

Although there is substantial agreement between CaT1 expressed in CHO cells and native  $I_{\text{CRAC}}$ , there are also differences. For example, on switching to divalent solution we found only a 5–10-fold increase in current, less than the approximate 10–40-fold increase reported for  $I_{\text{CRAC}}$ <sup>8,9</sup>. This difference depends, however, on the degree of  $I_{\text{CRAC}}$  or CaT1 activation at the time of the switch to DVF solution. Second, studies<sup>21</sup> have shown that CaT1, like  $I_{\text{CRAC}}$ <sup>2</sup>, has a much wider tissue distribution than reported originally<sup>4</sup>. Furthermore, we re-cloned CaT1 from RBL mast and Jurkat T lymphocyte cell lines, which suggests that CaT1, like  $I_{\text{CRAC}}$ , is present in blood cells. The failure of CaT1 detection in many cells is probably a result of the low levels of messenger RNA, corresponding to the paucity of  $I_{\text{CRAC}}$  channels in normal cells. It is also possible that the close CaT1 homologue, ECAC<sup>23</sup>, may account for  $I_{\text{CRAC}}$  in some cells.

We have shown that CaT1-expressing cells exhibit a  $\text{Ca}^{2+}$ -selective current with the properties of  $I_{\text{CRAC}}$ . CaT1 is activated by both passive and active store depletion, and this activation is not accounted for by the intrinsic CHO-K1  $I_{\text{CRAC}}$  current. We estimate that there are about 2,000 channels per cell in CaT1-expressing CHO-K1 cells, whereas there are less than 10 in wild-type CHO-K1 cells and about 100 in native resting lymphocytes<sup>8</sup>. Our data suggest that CaT1 is, or is a component of, the  $I_{\text{CRAC}}$  pore, but is dependent on the assumption that the single channel recordings reported in DVF solution<sup>8,9,13,16</sup> indeed represent  $I_{\text{CRAC}}$ . Future studies may use CaT1 to define the proteins that link it to intracellular  $\text{Ca}^{2+}$  stores and their co-expression may then recapitulate fully the properties of native  $I_{\text{CRAC}}$ . □

## Methods

### Electrophysiology and $\text{Ca}^{2+}$ measurements

The CaT1 clone from rat basophilic leukaemia (RBL) cells in the pTracerCMV2 (Invitrogen) vector was transiently transfected with *TransIT-LT1* (Pan Vera). Cells were transferred to coverslips 12 h after transfection and electrophysiological measurements ( $22 \pm 2^\circ\text{C}$ ) were made 24 h after transfection unless stated otherwise. The CaT1 or pTracerCMV2 vector-expressing cells were identified by green fluorescent protein fluorescence.

Whole-cell and single-channel currents were recorded using an Axopatch 200A amplifier. Data were digitized at 10 or 20 kHz, and digitally filtered off-line at 1 kHz. The internal pipette solution for macroscopic currents contained (in mM) 145 caesium methane sulphonate, 8 NaCl, 1  $\text{MgCl}_2$ , 10 HEPES, pH 7.2. We added 10 mM EGTA to this solution as a buffer for  $[\text{Ca}^{2+}]_i$  in most experiments. The standard extracellular solution contained (in mM): 145 NaCl, 5 CsCl, 1  $\text{MgCl}_2$ , 10 HEPES, 10 glucose, pH 7.4. We included 10 mM  $\text{Ca}^{2+}$  in the extracellular solution unless stated otherwise. We estimated that nominally  $[\text{Ca}^{2+}]$ -free solution contained  $10 \mu\text{M}$   $[\text{Ca}^{2+}]$ . Divalent free (DVF) solution contained (in mM) 115 NaCl, 10 HEPES, 10 sodium HEDTA and 10 glucose, with free  $[\text{Ca}^{2+}] \approx 1 \text{ nM}$ ; final  $[\text{Na}^+]_o = 145 \text{ mM}$ .  $\text{MgCl}_2$  was added and buffered to  $3 \mu\text{M}$  in some experiments (Fig. 2a, b). For single-channel measurements in cell-attached mode, high  $\text{K}^+$  DVF solution, containing (in mM) 120 potassium aspartic acid, 20 KCl, 1  $\text{MgCl}_2$ , 5 EGTA, 5 EDTA, 10 HEPES and 10 glucose, pH 7.2, was used as the extracellular solution to clamp the membrane potential to  $\sim 0 \text{ mV}$ . For both cell-attached and outside-out configurations, the pipette solution contained (in mM): 110 sodium methane sulphonate, 10 NaCl, 10 HEPES, 10 sodium HEDTA, pH 7.2. The same solution was used extracellularly in outside-out patches after adding 10 mM glucose. The temperature was  $22 \pm 2^\circ\text{C}$ .

The permeability of  $\text{Cs}^+$  was measured relative to that of  $\text{Na}^+$  in the absence of

extracellular  $\text{Ca}^{2+}$  ( $1 \text{ nM}$  free  $[\text{Ca}^{2+}]_o$ ) using the measured reversal potential according to the following equation<sup>25</sup>:

$$P_{\text{Ca}}/P_{\text{Na}} = ([\text{Na}^+]_o - [\text{Na}^+]_i \exp(E_{\text{rev}}/RT)) / ([\text{Cs}^+]_i \exp(E_{\text{rev}}/RT) - [\text{Cs}^+]_o) \quad (1)$$

where  $E_{\text{rev}}$  is the reversal potential,  $F$  is Faraday's constant,  $R$  the gas constant and  $T$  is absolute temperature. The permeability of the divalent cations  $\text{Ca}^{2+}$ ,  $\text{Ba}^{2+}$ ,  $\text{Sr}^{2+}$  and  $\text{Mn}^{2+}$ , relative to  $\text{Na}^+$ , was calculated using the reversal potential measured with 10 mM of the respective cation in the extracellular solution, using the equation:

$$P_x/P_{\text{Na}} = (1 + \exp(E_{\text{rev}}/RT)) \{ ([\text{Na}^+]_i + \alpha [\text{Cs}^+]_i) \exp(E_{\text{rev}}/RT) - [\text{Na}^+]_o - \alpha [\text{Cs}^+]_o \} / 4[X]_o \quad (2)$$

where  $P_x$  is the permeability of the respective divalent cation,  $[X]_o$  is its extracellular concentration, and  $\alpha = P_{\text{Cs}}/P_{\text{Na}}$ .

### $\text{Ca}^{2+}$ measurements

Cells were loaded with  $2 \mu\text{M}$  Fura-2 AM in the cell culture medium at  $37^\circ\text{C}$  for 30 min. We recorded Fura-2 ratios (F340/F380) on a MERLIN imaging system (Olympus).

### Data analysis

Group data are presented as mean  $\pm$  s.e.m. Statistical comparisons were made using analysis of variance and the  $t$ -test with Bonferroni correction. A two-tailed value of  $P < 0.05$  was taken to be statistically significant.

Received 12 January 2001; accepted 19 February 2001.

- Lewis, R. S. Store-operated calcium channels. *Adv. Second Messenger Phosphoprotein Res.* **33**, 279–307 (1999).
- Parekh, A. B. & Penner, R. Store depletion and calcium influx. *Physiol. Rev.* **77**, 901–930 (1997).
- Clapham, D. E. Calcium signaling. *Cell* **80**, 259–268 (1995).
- Peng, J. -B. *et al.* Molecular cloning and characterization of a channel-like transporter mediating intestinal calcium absorption. *J. Biol. Chem.* **274**, 22739–22746 (1999).
- Putney, J. W. Jr Muscarinic, alpha-adrenergic and peptide receptors regulate the same calcium influx sites in the parotid gland. *J. Physiol. (Lond.)* **268**, 139–149 (1977).
- Lewis, R. S. & Cahalan, M. D. Mitogen-induced oscillations of cytosolic  $\text{Ca}^{2+}$  and transmembrane  $\text{Ca}^{2+}$  current in human leukemic T cells. *Cell Regul.* **1**, 99–112 (1989).
- Hoth, M. & Penner, R. Depletion of intracellular calcium stores activates a calcium current in mast cells. *Nature* **355**, 353–355 (1992).
- Kerschbaum, H. H. & Cahalan, M. D. Single-channel recording of a store-operated  $\text{Ca}^{2+}$  channel in Jurkat T lymphocytes. *Science* **283**, 836–839 (1999).
- Kerschbaum, H. H. & Cahalan, M. D. Monovalent permeability, rectification, and ionic block of store-operated calcium channels in Jurkat T lymphocytes. *J. Gen. Physiol.* **111**, 521–537 (1998).
- Zweifach, A. & Lewis, R. S. Mitogen-regulated  $\text{Ca}^{2+}$  current of T lymphocytes is activated by depletion of intracellular  $\text{Ca}^{2+}$  stores. *Proc. Natl Acad. Sci. USA* **90**, 6295–6299 (1993).
- Hoth, M. & Penner, R. Calcium release-activated calcium current in rat mast cells. *J. Physiol.* **465**, 359–386 (1993).
- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer Associates, Sunderland, MA, 1992).
- Lepple-Wienhues, A. & Cahalan, M. D. Conductance and permeation of monovalent cations through depletion-activated  $\text{Ca}^{2+}$  channels ( $I_{\text{CRAC}}$ ) in Jurkat T cells. *Biophys. J.* **71**, 787–794 (1996).
- Fasolato, C. & Nilius, B. Store depletion triggers the calcium release-activated calcium current ( $I_{\text{CRAC}}$ ) in macrovascular endothelial cells—a comparison with Jurkat and embryonic kidney cell lines. *Pflügers Archiv* **436**, 69–74 (1998).
- Hoth, M. & Penner, R. Calcium release-activated calcium current in rat mast cells. *J. Physiol. (Lond.)* **465**, 359–86 (1993).
- Fomina, A. F., Fanger, C. M., Kozak, J. A. & Cahalan, M. D. Single channel properties and regulated expression of  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  (CRAC) channels in human T cells. *J. Cell. Biol.* **150**, 1435–44 (2000).
- Zweifach, A. & Lewis, R. S. Slow calcium-dependent inactivation of depletion-activated calcium current. Store-dependent and-independent mechanisms. *J. Biol. Chem.* **270**, 14445–14451 (1995).
- Montero, M., Alvarez, J. & Garcia-Sancho, J. Uptake of  $\text{Ca}^{2+}$  and refilling of intracellular  $\text{Ca}^{2+}$  stores in Ehrlich-ascites-tumour cells and in rat thymocytes. *Biochem. J.* **271**, 535–540 (1990).
- Philipp, S. *et al.* A novel capacitative calcium entry channel expressed in excitable cells. *EMBO J.* **17**, 4274–4282 (1998).
- Fierro, L. & Parekh, A. B. Substantial depletion of the intracellular  $\text{Ca}^{2+}$  stores is required for macroscopic activation of the  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  current in rat basophilic leukaemia cells. *J. Physiol. (Lond.)* **522**, 247–257 (2000).
- Peng, J. B. *et al.* Human calcium transport protein CaT1. *Biochem. Biophys. Res. Commun.* **278**, 326–332 (2000).
- Peng, J. -B. *et al.* Rat kidney-specific calcium transporter in the distal nephron. *J. Biol. Chem.* **275**, 28186–28194 (2000).
- Vennekens, R. *et al.* Permeation and gating properties of the novel epithelial  $\text{Ca}^{2+}$  channel. *J. Biol. Chem.* **275**, 3963–3969 (2000).

### Acknowledgements

We thank M. D. Cahalan for discussion.

Correspondence and requests for materials should be addressed to D.E.C. (e-mail: clapham@rascal.med.harvard.edu).

# IKK $\alpha$ controls formation of the epidermis independently of NF- $\kappa$ B

Yinling Hu\*, Veronique Baud\*†‡, Takefumi Oga\*†‡, Keun Il Kim\*, Kazuhiko Yoshida\*‡ & Michael Karin\*

\* Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0636, USA

† These authors contributed equally to this work

The IKK $\alpha$  and IKK $\beta$  catalytic subunits of I $\kappa$ B kinase (IKK) share 51% amino-acid identity and similar biochemical activities: they both phosphorylate I $\kappa$ B proteins at serines that trigger their degradation<sup>1–4</sup>. IKK $\alpha$  and IKK $\beta$  differ, however, in their physiological functions. IKK $\beta$  and the IKK $\gamma$ /NEMO regulatory subunit are required for activating NF- $\kappa$ B by pro-inflammatory stimuli and preventing apoptosis induced by tumour necrosis factor- $\alpha$  (refs 5–11). IKK $\alpha$  is dispensable for these functions, but is essential for developing the epidermis and its derivatives<sup>12–15</sup>. The mammalian epidermis is composed of the basal, spinous, granular and cornified layers<sup>16</sup>. Only basal keratinocytes can proliferate and give rise to differentiated derivatives, which on full maturation undergo enucleation to generate the cornified layer. Curiously, keratinocyte-specific inhibition of NF- $\kappa$ B, as in *Ikk $\alpha$ <sup>–/–</sup>* mice<sup>12–15</sup>, results in epidermal thickening but does not block terminal differentiation<sup>17,18</sup>. It has been proposed<sup>19,20</sup> that the epidermal defect in *Ikk $\alpha$ <sup>–/–</sup>* mice may be due to the failed activation of NF- $\kappa$ B. Here we show that the unique function of IKK $\alpha$  in control of keratinocyte differentiation is not exerted through its I $\kappa$ B kinase activity or through NF- $\kappa$ B. Instead, IKK $\alpha$  controls production of a soluble factor that induces keratinocyte differentiation.

The loss of IKK $\alpha$  prevents terminal differentiation of epidermal keratinocytes and increases their proliferative potential<sup>12,13</sup>. The *Ikk $\alpha$ <sup>–/–</sup>* epidermis exhibits increased expression of the proliferation markers cytokeratin (CK)-6 and proliferating cell nuclear antigen (PCNA)<sup>16,21</sup>, as well as CK5, CK1 and CK10, which are markers of basal and spinous cells<sup>16</sup>. By contrast, markers of terminal differentiation, loricrin and filaggrin<sup>16,21</sup>, are not expressed<sup>12–14</sup> (see Supplementary Information). To examine the cell autonomy of this defect, we established an *ex vivo* keratinocyte culture system. Both wild-type and *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes expressed CK5, but unlike wild-type cells that underwent terminal differentiation, as indicated by loricrin and filaggrin expression, *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes failed to do so even after 7 d (Fig. 1a).

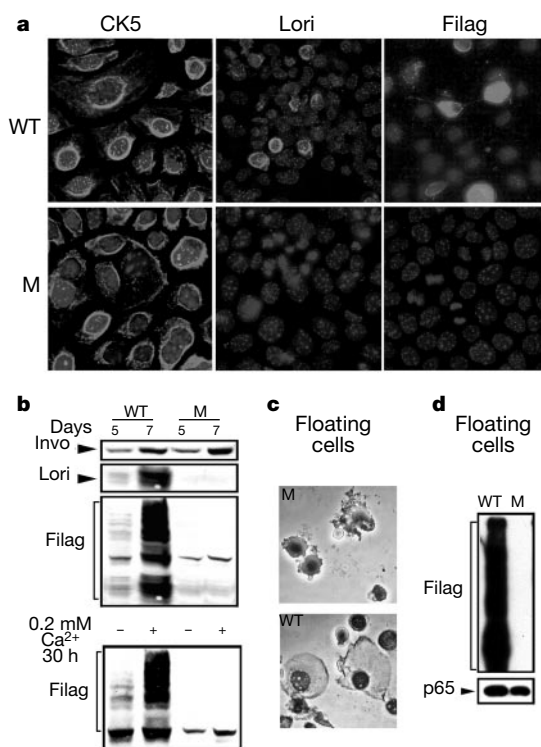
Similar results were obtained by immunoblot analysis, and even a 30-h culture in the presence of elevated Ca<sup>2+</sup> failed to induce differentiation of *Ikk $\alpha$ <sup>–/–</sup>* cells (Fig. 1b). As cultured keratinocytes differentiate, they enlarge, round up and eventually detach from the substratum<sup>22</sup>. Large floating cells that expressed filaggrin were readily detected in 7-d wild-type but not *Ikk $\alpha$ <sup>–/–</sup>* cultures (Fig. 1c, d). Cultured *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes also exhibited the hyperproliferative phenotype seen *in vivo*, and even 0.5 mM Ca<sup>2+</sup> did not inhibit their proliferation (see Supplementary Information).

It has been suggested that the loss of IKK $\alpha$  results in defective activation of NF- $\kappa$ B in the epidermis<sup>13,14</sup>. We re-examined the state of the NF- $\kappa$ B pathway in wild-type and *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes. Instead of decreased NF- $\kappa$ B activity, *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes exhibited higher than normal levels of IKK and NF- $\kappa$ B after stimulation

with either tumour necrosis factor (TNF)- $\alpha$  or interleukin-1 (see Supplementary Information). This is probably due to the replacement of IKK $\alpha$ –IKK $\beta$  heterodimers with more active IKK $\beta$ –IKK $\beta$  homodimers. Thus, *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes do not exhibit a primary NF- $\kappa$ B activation defect.

To pinpoint the relationship between IKK $\alpha$  and NF- $\kappa$ B during keratinocyte differentiation, we infected wild-type and mutant keratinocytes with adenoviruses expressing IKK $\alpha$ , IKK $\beta$ , p65 (RelA) and the degradation-resistant I $\kappa$ B $\alpha$ (A32/36) mutant. If p65 is a downstream effector of IKK $\alpha$  in this pathway, its overexpression should induce differentiation of *Ikk $\alpha$ <sup>–/–</sup>* cells. Likewise, I $\kappa$ B $\alpha$ (A32/36) should block differentiation of wild-type cells. A similar effect is expected from catalytically inactive IKK $\beta$ (KA), which inhibits NF- $\kappa$ B activation<sup>4</sup>. In addition, IKK $\beta$  may substitute for IKK $\alpha$  if the target for IKK $\alpha$  is NF- $\kappa$ B. Unexpectedly, only IKK $\alpha$  led to the appearance of giant cells when introduced into *Ikk $\alpha$ <sup>–/–</sup>* cultures, suggesting that differentiation was restored (Fig. 2a). Expression of IKK $\beta$ , IKK $\beta$ (KA), p65 or I $\kappa$ B $\alpha$ (A32/36) in *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes did not elicit such an effect or other morphological changes (Fig. 2a; and data not shown). Furthermore, none of these vectors blocked the induction of giant cells when co-infected with IKK $\alpha$  (data not shown).

Most notably, infection of *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes with IKK $\alpha$  induced filaggrin expression (Fig. 2b). The other viruses did not



**Figure 1** Defective differentiation of cultured *Ikk $\alpha$ <sup>–/–</sup>* keratinocytes. **a**, After 3–7 d in culture, wild-type (WT) and *Ikk $\alpha$ <sup>–/–</sup>* (M) keratinocytes were fixed and stained with antibodies to CK5 (at day 3), loricrin (lori) or filaggrin (filag; at day 7). Nuclei were counterstained with DAPI. Cells were examined by indirect immunofluorescence. **b**, Lysates of cultured *Ikk $\alpha$ <sup>+/+</sup>* (WT) and *Ikk $\alpha$ <sup>–/–</sup>* (M) keratinocytes were examined for expression of involucrin (invo), loricrin and filaggrin by immunoblotting. Top, cells kept for 5 or 7 d in normal medium. Bottom, cells kept for 30 h without or with 0.2 mM Ca<sup>2+</sup> after 5 d in culture. **c**, Morphology of floating cells collected from WT and *Ikk $\alpha$ <sup>–/–</sup>* cultures, stained with haematoxylin and eosin. **d**, Filaggrin expression in lysates of floating WT and *Ikk $\alpha$ <sup>–/–</sup>* cells. p65 was used as a loading control. Because of extensive processing, profilaggrin and filaggrin appear as multiple species with different electrophoretic mobilities.

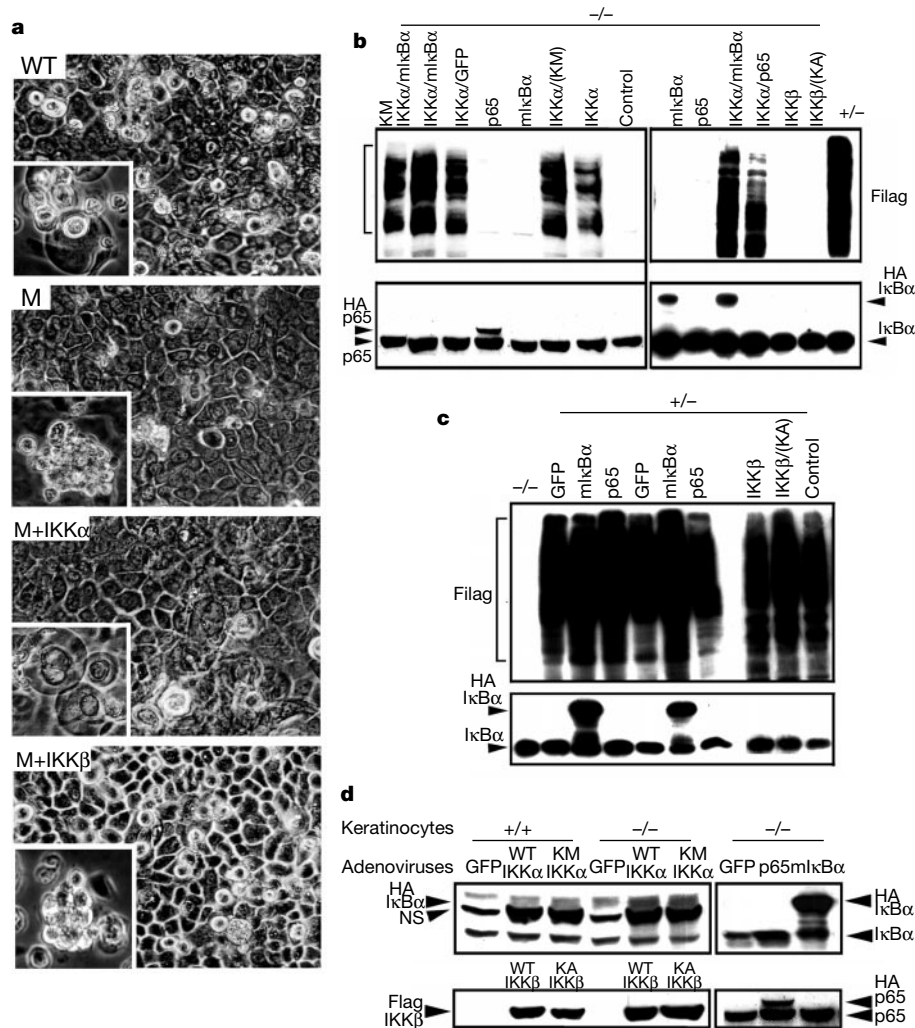
‡ Present addresses: Laboratoire Oncogène, Différenciation et Transduction du Signal, Institut André Lwoff, 7, Rue Guy Moquet, 94800 Villejuif, France (V.B.); Department of Surgery II, Faculty of Medicine, Kyushu University, 3-1-1 Maidashi, Higashiku, Fukuoka 812-8582, Japan (T.O.); Department of Ophthalmology, Hokkaido University, N15 W7, Kita-ku Sapporo 060-8638, Japan (K.Y.).

exhibit such an activity; neither did they inhibit IKK $\alpha$ -induced differentiation. Furthermore, neither p65, I $\kappa$ B $\alpha$ (A32/36), IKK $\beta$  nor IKK $\beta$ (KA) inhibited the differentiation of *Ikk $\alpha$ <sup>+/-</sup>* cells (Fig. 2c). As expected<sup>17</sup>, I $\kappa$ B $\alpha$ (A32/36) enhanced, and p65 attenuated, the proliferation of cultured keratinocytes (data not shown). All the relevant proteins were expressed in both wild-type and *Ikk $\alpha$ <sup>-/-</sup>* keratinocytes (Fig. 2d) and were functional, because I $\kappa$ B $\alpha$ (A32/36) inhibited NF- $\kappa$ B activity (see below), IKK $\beta$  generated high I $\kappa$ B kinase activity, and IKK $\beta$ (KA) inhibited IKK activity (data not shown). These results suggest that IKK $\alpha$  has a unique role in inducing keratinocyte differentiation, which is not mediated through NF- $\kappa$ B.

We next constructed several *Ikk $\alpha$*  mutants and tested their differentiation-inducing activity. The mutants included two carboxy-terminal truncations, C40 (residues 1–705) and C80 (1–665), and three kinase domain mutants, IKK $\alpha$ (KM), IKK $\alpha$ (AA) and IKK $\alpha$ (EE), which either destroy the ATP-binding site of IKK $\alpha$  or convert its activating phosphoacceptor sites to alanines or glutamates, respectively<sup>3,23</sup>. We also tested the activities of mutants with defective leucine zipper (LZ<sup>-</sup>) or helix–loop–helix (HLH<sup>-</sup>) motifs<sup>4</sup>, as well as an amino-terminal deletion of residues 1–390 (Fig. 3a).

Only wild-type IKK $\alpha$ , IKK $\alpha$ (EE), C40 and C80 exhibited I $\kappa$ B kinase activity when overexpressed in HEK293 cells (Fig. 3b, row 2). Except for C40 and C80, all of the IKK $\alpha$  mutants interacted with IKK $\gamma$  (Fig. 3b, row 4). The activities of the different mutants were also tested in keratinocytes and found to be the same as in HEK293 cells (data not shown). In agreement with previous experiments<sup>4,23,24</sup>, the I $\kappa$ B kinase activity of the overexpressed IKK $\alpha$  proteins was constitutive. At the levels expressed, none of the IKK $\alpha$  mutants inhibited endogenous IKK or NF- $\kappa$ B (Fig. 3b, rows 3 and 5). Unexpectedly, the kinase activity of IKK $\alpha$  was not required for induction of filaggrin expression (Fig. 3c). Furthermore, this activity did not require direct binding to IKK $\gamma$ , as C40 and C80 were as active (after normalization for expression levels) as wild-type IKK $\alpha$  (Fig. 3c). The LZ<sup>-</sup>, HLH<sup>-</sup> and  $\Delta$ 1–390 mutants, however, were inactive or only partially capable of inducing differentiation of *Ikk $\alpha$ <sup>-/-</sup>* keratinocytes (Fig. 3c). Thus, induction of keratinocyte differentiation does not require the protein kinase activity of IKK $\alpha$ . All of the mutants were expressed efficiently (Fig. 3d).

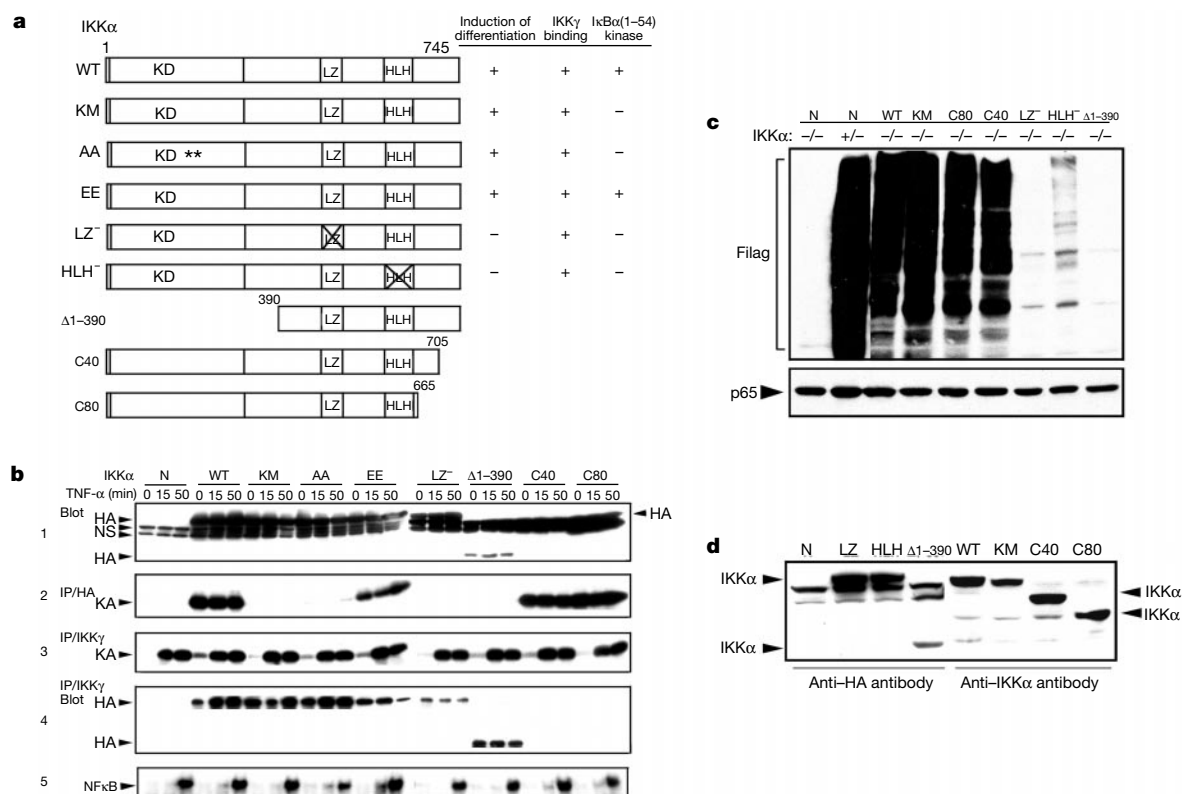
To characterize the epidermal differentiation defect, we transplanted *Ikk $\alpha$ <sup>-/-</sup>* skin from embryonic day (E)17–18 fetuses to



**Figure 2** Introduction of IKK $\alpha$  into *Ikk $\alpha$ <sup>-/-</sup>* keratinocytes induces their differentiation. **a**, *Ikk $\alpha$ <sup>-/-</sup>* (M) and WT keratinocytes were infected at day 3 of culture with adenoviruses encoding IKK $\alpha$  and IKK $\beta$ , either alone or in combination. The cultures were photographed 5 d after infection; original magnification,  $\times 200$ . Floating cells are shown at the bottom left insets. **b**, **c**, Filaggrin expression in *Ikk $\alpha$ <sup>-/-</sup>* and WT (+/-) cells infected with the indicated

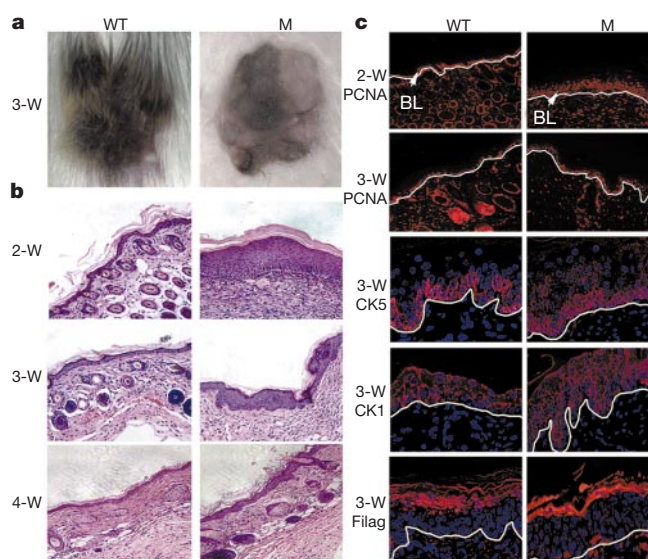
adenoviruses. p65 and I $\kappa$ B $\alpha$  expression was also examined. The cells were infected on day 3 of culture and analysed 5 d later. **d**, Expression of ectopic IKK $\alpha$ , IKK $\beta$ , p65 and I $\kappa$ B $\alpha$ (A32/36) in adenovirus-infected keratinocytes. The haemagglutinin A (HA) antibody detects nonspecific (NS) crossreactive polypeptides, one of which migrates close to HA-IKK $\alpha$ .





**Figure 3** IKKα kinase activity or interaction with IKKγ is not required for induction of keratinocyte differentiation. **a**, Summary of IκB kinase activities, IKKγ binding, and differentiation-inducing ability of the different IKKα constructs. HLH, helix-loop-helix; KD, kinase domain; LZ, leucine zipper. Asterisks indicate sites of amino-acid substitutions. **b**, Analysis of IKK activity and IKKγ binding. IKKα constructs were expressed using recombinant adenoviruses in HEK293 cells. N, adenovirus with a GFP insert. Row 1, expression of HA-tagged proteins. Rows 2 and 3, IκB kinase activity (KA) of proteins immunoprecipitated with anti-HA or anti-IKKγ antibody. Row 4, IKKα

polypeptides binding to IKKγ examined by anti-HA immunoblot analysis of anti-IKKγ isolated complexes. Row 5, NF-κB DNA-binding activity analysed by gel mobility shift assay. In all cases, cells infected with IKKα viruses or a control (N) virus were treated with TNF-α for the indicated times 48 h after infection. **c**, Filaggrin expression by *Ikkα*<sup>-/-</sup> keratinocytes (-/-) infected with the indicated adenoviruses was analysed by immunoblotting 5–7 days after infection. *Ikkα*<sup>+/-</sup> cells (+/-) were used as a positive control. Levels of p65 were used to control loading. **d**, Expression of ectopic IKKα proteins in *Ikkα*<sup>-/-</sup> keratinocytes.

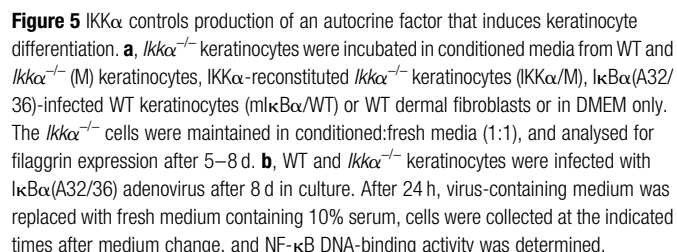


**Figure 4** Normal differentiation of engrafted *Ikkα*<sup>-/-</sup> skin. **a**, Formation of hair by WT and *Ikkα*<sup>-/-</sup> (M) (C57Bl/6xSv129 background) skin grafts 3 weeks (3-W) after transplantation to a *Rag1*<sup>-/-</sup> host (FVB/129 background). **b**, Histological appearance of WT and *Ikkα*<sup>-/-</sup> skin grafts at 2, 3 and 4 weeks after transplantation onto a *Rag1*<sup>-/-</sup> host. Paraffin-embedded sections were stained with haematoxylin and eosin. Original magnification,

×200. **c**, Expression of proliferation and differentiation markers by WT and *Ikkα*<sup>-/-</sup> skin grafts 2 and 3 weeks after transplantation. Paraffin-embedded skin sections were stained with antibodies to PCNA, CK5, CK1 or filaggrin, and examined by indirect immunofluorescence. Original magnification, ×200. White line indicates the border between epidermis and dermis. BL, basal layer.

To test this hypothesis, we collected conditioned media from 8–12-day-old cultures of wild-type and *Ikkα*<sup>−/−</sup> keratinocytes, IKKα reconstituted *Ikkα*<sup>−/−</sup> keratinocytes and wild-type dermal fibroblasts, and examined their ability to induce differentiation of *Ikkα*<sup>−/−</sup> keratinocytes. Only media conditioned by wild-type or IKKα-reconstituted *Ikkα*<sup>−/−</sup> keratinocytes induced filaggrin expression (Fig. 5a). Such an activity was not present in media conditioned by *Ikkα*<sup>−/−</sup> keratinocytes or wild-type dermal fibroblasts. To determine whether production of this keratinocyte differentiation-inducing factor (kDIF) depends on NF-κB, we infected wild-type

These experiments define the primary cellular defect caused by the loss of the IKK $\alpha$  subunit of the IKK complex as a failure to activate the terminal differentiation program of epidermal keratinocytes. This program includes permanent growth arrest, expression of structural proteins, such as loricrin and filaggrin, and cell death through enucleation<sup>16</sup>. As a result, *Ikka*<sup>-/-</sup> keratinocytes continue to proliferate even in the presence of high extracellular Ca<sup>2+</sup>, a potent inducer of growth arrest and terminal differentiation<sup>26</sup>. Terminal differentiation of *Ikka*<sup>-/-</sup> keratinocytes can be rescued completely by adenovirus-mediated transduction of a wild-type IKK $\alpha$  complementary DNA, but in contrast to previous expectations<sup>13,19,20</sup> the differentiation-inducing ability of IKK $\alpha$  is not mediated through NF- $\kappa$ B. In addition, it does not depend on its protein kinase activity or direct binding to the IKK $\gamma$ /NEMO subunit—an interaction that is essential for IKK and NF- $\kappa$ B activation<sup>7,8</sup>. Nonetheless, these findings are consistent with the



**c.** Characterization of the differentiation-inducing factor. Conditioned medium collected from WT keratinocytes was heated at 100 °C for 5–60 min, digested with 20  $\mu\text{g ml}^{-1}$  native (Try) or heat-inactivated (H-try) trypsin for 2 h at 37 °C. Conditioned or normal (N) media were added to *IKK $\alpha$ <sup>-/-</sup>* keratinocytes, analysed as in **a**, **c**, **L**, loading control. **d**, IKK $\alpha$  is a bifunctional protein. In the IKK complex it contributes to I $\kappa$ B phosphorylation and NF- $\kappa$ B activation; however, this function is not essential and can be replaced by IKK $\beta$ . The essential function of IKK $\alpha$  in epidermal differentiation, which is mediated by a keratinocyte differentiation-inducing factor (kDIF), is likely to be exerted by IKK $\alpha$  that is not associated with other IKK components.

full differentiation ability of *Ikkβ*- or *Ikkγ/Nemo*-deficient keratinocytes<sup>7,8,10</sup>. Thus, the most essential and unique function of IKKα in keratinocyte differentiation, a function that cannot be replaced by IKKβ, is not mediated by the classical IKK complex. Although it remains unknown whether the differentiation-inducing activity of IKKα requires association with another protein, IKKα is clearly a bifunctional protein (Fig. 5d). As a component of the IKK complex it can contribute to IKK and NF-κB activation, but this function is not essential or unique and is fully replaced by IKKβ<sup>24,27</sup>. Independently of the IKK complex, IKKα exerts its essential and most unique function—induction of keratinocyte differentiation.

Instead of acting through IKK and NF-κB, IKKα controls the production of a differentiation-inducing factor, kDIF, that is essential for proper differentiation of the mammalian epidermis. On the basis of the *Ikkα*<sup>-/-</sup> phenotype, kDIF is different from other polypeptide factors, such as transforming growth factor (TGF)-α, TGF-β or other growth factors that may contribute to formation of the epidermis<sup>28</sup>. kDIF is not produced by dermal fibroblasts, but it is possible that its *in vivo* production is stimulated by an interaction of the epidermis with the dermis, a classical epithelio-mesenchymal interaction required for skin differentiation. Although the molecular identity of kDIF is not yet known, we suspect that the *pupoid fetus* (*Pf*) and *repeated epilation* (*Er*) mutations, whose phenotypes are identical to that of the *Ikkα*<sup>-/-</sup> deficiency<sup>29</sup>, affect either its production or activity and may even map to the gene coding for kDIF. Both the *Pf* and *Er* mutant alleles were mapped to mouse chromosome 4 (ref. 29), whereas the *Ikkα* gene is on mouse chromosome 19 (ref. 30). *Pf* and *Er* are therefore not allelic to *Ikkα*<sup>-/-</sup>, and represent mutations in a different genetic locus whose product is also a component of the signalling cascade controlled by IKKα in the epidermis. □

## Methods

### Isolation and culture of mouse keratinocytes

Skins of E18 to newborn mice were removed and placed in serum-free Defined Keratinocyte-SFM medium (GibcoBRL) at 4 °C overnight. After genotyping, the skins were incubated in 0.25% trypsin (GibcoBRL) at 4 °C for 8–10 h. We removed the dermis and transferred the epidermis into tubes containing culture medium. Keratinocytes from the inner surface of the epidermis were isolated by shaking and filtering through cell strainers (Fisher). We used 2 × 10<sup>5</sup> cells per 6-cm dish in colony formation and cell growth assays, and (5–10) × 10<sup>5</sup> cells per 6-cm dish in differentiation experiments. To count colonies, cultures were washed with PBS once, fixed with 100% ethanol at -20 °C, stained with 0.1% crystal blue for 40 min, and washed with water for three times. Cell morphology was determined by microscopical examination of non-fixed cultures.

### Adenovirus constructs and infection

Wild-type and mutant human IKKα cDNA fragments, including IKKα(KM), Δ1–390(residues 390–745), IKKα-HLH<sup>+</sup>, IKKα-LZ<sup>-</sup>, IKKα(AA) and IKKα(EE), as well as wild-type IKKβ and IKKβ(KA) cDNAs were excised from described expression vectors<sup>3,4,23</sup>, and subcloned between the *Hind*III and *Nor*I sites of the pXCJL1 (Ad5) vector (Microbix Biosystems). The truncated IKKα C40 (residues 1–705) and C80 (1–665) cDNAs were generated by polymerase chain reaction and subcloned as above. Details of all constructs are available on request. Recombinant adenoviruses that were plaque purified were generated at the Vector Core Development Laboratory of UCSD. IKKα(A32/36), p65 and green fluorescent protein (GFP) adenoviruses were kindly provided by D. Brenner (Univ. North Carolina) and B. Benett (Signal Pharmaceuticals, San Diego). We infected cells with 1 × 10<sup>8</sup> p.f.u. adenovirus after day 3 of culturing. Virus-containing media were replaced by fresh media 1 d after infection.

### Immunotechniques, and IKK and NF-κB assays

Paraffin sections of mouse skin grafts were immunostained with antibodies to CK5 (Babco), CK1 (Babco), PCNA (Pharmingen), loricrin (Babco) or flaggrin (Babco), and examined by indirect immunofluorescence as described<sup>15</sup>. Immunoblot analysis and immunocomplex kinase assays of keratinocyte lysates were done as described<sup>1</sup>. NF-κB DNA-binding activity was measured by a gel mobility shift assay<sup>1</sup>.

### Skin grafts

We surgically transplanted *Ikkα*<sup>+/-</sup> and *Ikkα*<sup>-/-</sup> mouse embryo (E18–19) skins into the dorsal region of 6–8-week-old anaesthetized *Rag1*<sup>-/-</sup> mice (FVB/129 strain). The average size of the donor graft was 1–1.5 cm in diameter. The grafts were dressed with a trimmed 3M Tegaderm bandage. After 4–5 d, the bandages were removed, and the grafts were

examined 1–4 weeks after surgery. For *Ikkα*<sup>-/-</sup> grafts the donor origin was verified by β-galactosidase staining.

Received 1 November 2000; accepted 24 January 2001.

- DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive IκB kinase that activates the transcription factor NF-κB. *Nature* **388**, 548–554 (1997).
- Regnier, C. H. et al. Identification and characterization of an IκB kinase. *Cell* **90**, 373–383 (1997).
- Mercurio, F. et al. IKK-1 and IKK-2: cytokine-activated IκB kinases essential for NF-κB activation. *Science* **278**, 860–866 (1997).
- Zandi, E., Rothwarf, D. M., Delhase, M., Hayakawa, M. & Karin, M. The IκB kinase complex (IKK) contains two kinase subunits, IKKα and IKKβ, necessary for IκB phosphorylation and NF-κB activation. *Cell* **91**, 243–252 (1997).
- Rothwarf, D. M., Zandi, E., Natoli, G. & Karin, M. IKKγ is an essential regulatory subunit of the IκB kinase complex. *Nature* **395**, 297–300 (1998).
- Yamaoka, S. et al. Complementation cloning of NEMO, a component of the IκB kinase complex essential for NF-κB activation. *Cell* **93**, 1231–1240 (1998).
- Makris, C. et al. Female mice heterozygous for IKKγ/NEMO deficiencies develop a dermatopathy similar to the human X-linked disorder incontinentia pigmenti. *Mol. Cell* **5**, 969–979 (2000).
- Schmidt-Suppran, M. et al. NEMO/IKKγ-deficient mice model incontinentia pigmenti. *Mol. Cell* **5**, 981–992 (2000).
- Li, Q., Van Antwerp, D., Mercurio, F., Lee, K. F. & Verma, I. M. Severe liver degeneration in mice lacking the IκB kinase 2 gene. *Science* **284**, 321–325 (1999).
- Li, Z. W. et al. The IKKβ subunit of IκB kinase (IKK) is essential for NF-κB activation and prevention of apoptosis. *J. Exp. Med.* **189**, 1839–1845 (1999).
- Tanaka, M. et al. Embryonic lethality, liver degeneration, and impaired NF-κB activation in IKKβ-deficient mice. *Immunity* **10**, 421–429 (1999).
- Hu, Y. et al. Abnormal morphogenesis but intact IKK activation in mice lacking the IKKα subunit of IκB kinase. *Science* **284**, 316–320 (1999).
- Takeda, K. et al. Limb and skin abnormalities in mice lacking IKKα. *Science* **284**, 313–316 (1999).
- Li, Q. et al. IKK1-deficient mice exhibit abnormal development of skin and skeleton. *Genes Dev.* **13**, 1322–1328 (1999).
- Yoshida, K., Hu, Y. & Karin, M. IκB kinase alpha is essential for development of the mammalian cornea and conjunctiva. *Invest. Ophthalmol. Vis. Sci.* **41**, 3665–3669 (2000).
- Fuchs, E. & Byrne, C. The epidermis: rising to the surface. *Curr. Opin. Genet. Dev.* **4**, 725–736 (1994).
- Seitz, C. S., Lin, Q., Deng, H. & Khavari, P. A. Alterations in NF-κB function in transgenic epithelial tissue demonstrate a growth inhibitory role for NF-κB. *Proc. Natl Acad. Sci. USA* **95**, 2307–2312 (1998).
- van Hogerlinden, M., Rozell, B. L., Ahrlund-Richter, L. & Toftgard, R. Squamous cell carcinomas and increased apoptosis in skin with inhibited Rel/NF-κB signaling. *Cancer Res.* **59**, 3299–3303 (1999).
- Seitz, C. S., Freiberg, R. A., Hinata, K. & Khavari, P. A. NF-κB determines localization and features of cell death in epidermis. *J. Clin. Invest.* **105**, 253–260 (2000).
- May, M. J. & Ghosh, S. IκB kinases: kinsmen with different crafts. *Science* **284**, 271–273 (1999).
- Knuchel, R., Hofstaedter, F., Sutherland, R. M. & Keng, P. C. Proliferation-associated antigens PCNA and Ki-67 in two- and three-dimensional experimental systems of human squamous epithelial carcinomas. *Verh. Dtsch. Ges. Pathol.* **74**, 275–278 (1990).
- Sun, T. T. & Green, H. Differentiation of the epidermal keratinocyte in cell culture: formation of the cornified envelope. *Cell* **9**, 511–521 (1976).
- Delhase, M., Hayakawa, M., Chen, Y. & Karin, M. Positive and negative regulation of IκB kinase activity through IKKβ subunit phosphorylation. *Science* **284**, 309–313 (1999).
- Rothwarf, D. M. & Karin, M. The NF-κB activation pathway: a paradigm in information transfer from membrane to nucleus. *Science STKE* [online] (cited 26 October 1999) ([www.stke.org/cgi/content/full/OC\\_sigrans;1999/5/rel](http://www.stke.org/cgi/content/full/OC_sigrans;1999/5/rel)) (1999).
- Oettinger, M. A. Activation of V(D)J recombination by RAG1 and RAG2. *Trends Genet.* **8**, 413–416 (1992).
- Yuspa, S. H. & Morgan, D. L. Mouse skin cells resistant to terminal differentiation associated with initiation of carcinogenesis. *Nature* **293**, 72–74 (1981).
- Li, Q., Estepa, G., Memet, S., Israel, A. & Verma, I. M. Complete lack of NF-κB activity in IKK1 and IKK2 double-deficient mice: additional defect in neurulation. *Genes Dev.* **14**, 1729–1733 (2000).
- Tomic-Canic, M., Komine, M., Freedberg, I. M. & Blumenberg, M. Epidermal signal transduction and transcription factor activation in activated keratinocytes. *J. Dermatol. Sci.* **17**, 167–181 (1998).
- Fisher, C., Jones, A. & Roop, D. R. Abnormal expression and processing of keratins in pupoid fetus (*pf*) and repeated epilation (*Er/Er*) mutant mice. *J. Cell Biol.* **105**, 1807–1819 (1987).
- Mock, B. A., Connelly, M. A., McBride, O. W., Kozak, C. A. & Marcu, K. B. CHUK, a conserved helix-loop-helix ubiquitous kinase, maps to human chromosome 10 and mouse chromosome 19. *Genomics* **27**, 348–351 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

### Acknowledgements

We thank K. M. Hodivala-Dilke for advice on keratinocyte isolation; K. Bouic for assistance with adenovirus preparation; D. Brenner and B. Benett for recombinant adenoviruses; M. Delhase for IKKα (EE and 390) mutants; and J. Feramisco for assistance with deconvolution microscopy at the UCSD Cancer Center Core Facility. Y.H., V.B. and T.O. were supported by postdoctoral fellowships from the Arthritis Foundation, Human Frontier Science Program and Japanese Society for Promotion of Science, respectively. Work was supported by grants from the NIH, the Association For International Cancer Research and the CERES Research Award to M.K., who is the Frank and Else Schilling American Cancer Society Research Professor.

Correspondence and requests for materials should be addressed to M.K. (e-mail: [karinoffice@ucsd.edu](mailto:karinoffice@ucsd.edu)).



# Functional proteins from a random-sequence library

Anthony D. Keefe & Jack W. Szostak

Howard Hughes Medical Institute, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Functional primordial proteins presumably originated from random sequences, but it is not known how frequently functional, or even folded, proteins occur in collections of random sequences. Here we have used *in vitro* selection of messenger RNA displayed proteins, in which each protein is covalently linked through its carboxy terminus to the 3' end of its encoding mRNA<sup>1</sup>, to sample a large number of distinct random sequences. Starting from a library of  $6 \times 10^{12}$  proteins each containing 80 contiguous random amino acids, we selected functional proteins by enriching for those that bind to ATP. This selection yielded four new ATP-binding proteins that appear to be unrelated to each other or to anything found in the current databases of biological proteins. The frequency of occurrence of functional proteins in random-sequence libraries appears to be similar to that observed for equivalent RNA libraries<sup>2,3</sup>.

The frequency of occurrence of functional proteins in collections of random sequences is an important constraint on models of the evolution of biological proteins. Here we have experimentally determined this frequency by isolating proteins with a specific function from a large random-sequence library of known size. We selected for proteins that could bind a small molecule target with high affinity and specificity as a way of identifying amino-acid sequences that could form a three-dimensional folded state with a well-defined binding site and therefore exhibit an arbitrary specific function. ATP was chosen as the target for binding to allow comparison with known biological ATP-binding motifs and also with previous selections using random-sequence RNA libraries<sup>2,3</sup>.

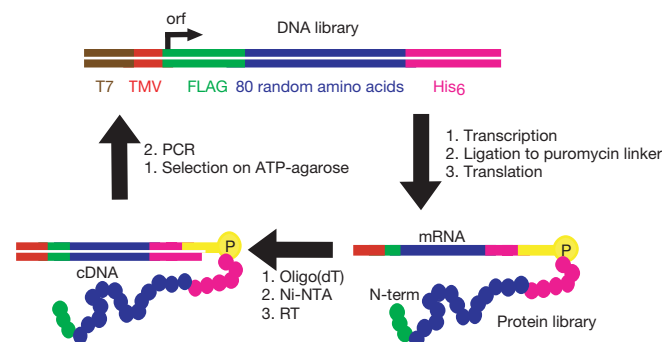
Because protein sequences with specific functions are expected to be quite rare in protein sequence space, we prepared a DNA library of  $4 \times 10^{14}$  independently generated random sequences. This DNA library was specifically constructed to avoid stop codons and frameshift mutations<sup>4</sup>, and was designed for use in mRNA display<sup>1</sup> selections. This DNA library was then used to generate  $6 \times 10^{12}$

purified non-redundant random proteins that were used as the input into the first selection step. These proteins contain a contiguous stretch of random amino acids 80 residues in length, long enough to form known protein domains. Unlike other libraries that have been used in protein selections, this random region is not part of a larger structure that would otherwise tend to constrain or bias the conformation of the resulting proteins. This library randomly samples the whole of sequence space, rather than the vicinity of a known protein. The random region of each library member is flanked by short invariant sequences encoding affinity tags for purification (Fig. 1).

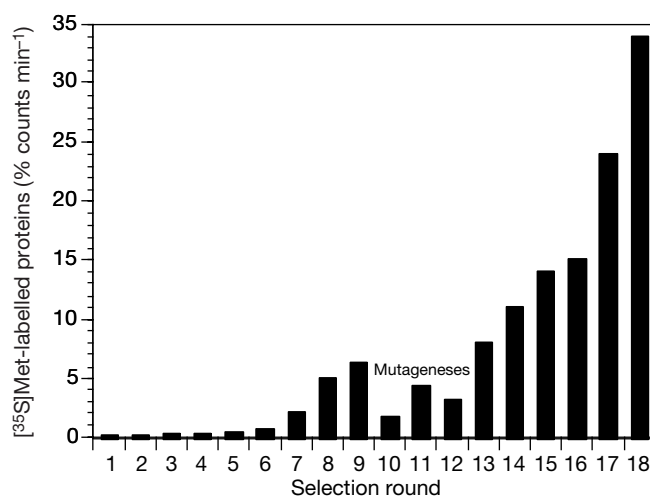
Successive rounds of *in vitro* selection and amplification were performed starting with this random-sequence library. In each round the mRNA-displayed proteins were incubated with immobilized ATP, washed and eluted with free ATP. The eluted fractions were collected and amplified by polymerase chain reaction (PCR); this DNA was then used to generate a new library of mRNA-displayed proteins, enriched in sequences that bind ATP, for input into the next round of selection (Fig. 1). After eight rounds, the fraction of mRNA-displayed proteins eluting with ATP had risen from 0.1 to 6.2% (Fig. 2). We cloned and sequenced 24 individual library members, which showed that the population was now dominated by 4 families of ATP-binding proteins (Fig. 3a). These families show no sequence relationship to each other or to any known biological protein. The members of each family are closely related, indicating that each family is descended from a single ancestral molecule, which was one of the original random sequences.

Single representatives of each of these protein families (round 8) were chosen for further study. Only 5–15% of the mRNA-displayed protein prepared from each of these clones binds to immobilized ATP and then elutes with free ATP under selection conditions, consistent with the 6.2% binding and elution with ATP for the library as a whole. One possible explanation for this low level of ATP-binding is conformational heterogeneity, possibly reflecting inefficient folding of these primordial protein sequences.

In an effort to increase the proportion of these proteins that fold into an ATP-binding conformation, we mutagenized the library and carried out further rounds of *in vitro* selection and amplification. Three consecutive rounds with mutagenic PCR amplification were performed with an average mutagenic rate of 3.7% per amino acid for each round. After six subsequent rounds of *in vitro* selection and amplification without mutagenesis, the proportion of the library of mRNA-displayed proteins that bound and eluted with ATP rose to



**Figure 1** *In vitro* selection and amplification of mRNA-displayed proteins. The DNA library encodes proteins with 80 contiguous random amino acids<sup>4</sup>. Transcription, splinted ligation to a 3'-puromycin oligonucleotide, translation, high-salt incubation, purification and reverse transcription (RT) yielded  $6 \times 10^{12}$  independent mRNA-displayed proteins. This library was incubated with an ATP-agarose affinity matrix. Unbound material was washed away with selection buffer at 4 °C, and bound material was collected by incubation with the same buffer containing 5 mM ATP at 4 °C. Eluted fractions were combined, amplified by PCR, and used as the input for the next round of *in vitro* selection and amplification.



**Figure 2** Proportion of the mRNA-displayed protein library bound to immobilized ATP and subsequently eluted with free ATP, as a function of selection round. The inputs into rounds 10–12 were subjected to mutagenic PCR amplification with an average mutagenesis rate of 3.7% at the amino-acid level per amplification.

34% (round 18) (Fig. 2). At this point the library was entirely composed (56/56 clones sequenced) of the descendents of one of the four originally selected protein families (family B) (Fig. 3b).

Comparing the round 18 sequences with the ancestral sequence showed that four amino-acid substitutions had become predominant in the selected population (present more than 39 times in 56 sequences, Fig. 3b), and that 16 other substitutions had also been selectively enriched (present more than 4 times in 56 sequences, Fig. 3b). In addition, each clone contained a variable number of other substitutions. The selectively enriched substitutions are distributed over the 62 amino-terminal amino acids of the original 80-amino-acid random region, suggesting that amino acids throughout this region are contributing to the formation of a folded structure, at least in the complex with ATP. The substitutions in each of the assayed clones improve ATP-binding relative to the ancestral sequence.

Eight individual proteins from the final round of *in vitro* selection (round 18) were chosen randomly for further study. As free proteins, the proportion that bound to the column and then eluted with ATP varied from 5% to 40%. The four proteins that bound and eluted to the greatest extent were expressed in *Escherichia coli* as C-terminal fusion proteins of maltose binding protein (MBP), as the solubilities of the free proteins were too low to permit full characterization. Purification on an amylose column and subsequently on a Ni-NTA column under denaturing conditions yielded several milligrams of each as highly pure fusion proteins.

The four MBP fusion proteins have dissociation constants ( $K_d$ ) for ATP that fall within the range 100 nM to 10  $\mu$ M at 4 °C, as determined by equilibrium dialysis with [ $\alpha$ - $^{32}$ P]ATP. Of these, protein 18-19 binds most strongly, with a  $K_d$  of 100 nM for ATP

at 4 °C and a 1:1 stoichiometry. Gel filtration indicates that 65% of this protein is monomeric in selection binding buffer, with the remainder as high-order aggregates running at the void volume.

Only the monomer binds ATP, and no low-order aggregates such as dimers or tetramers were observed. Competition experiments with ATP analogues show that protein 18-19 interacts with several distinct parts of the ATP molecule (Fig. 4). Removing one, two or three phosphate groups successively raises the  $K_d$ , with a particularly large increase accompanying the removal of the  $\alpha$ -phosphate (70-fold). Removing the ribose 2' or 3' hydroxyl groups reduces binding by factors of 40 and 3, respectively. Binding to ITP (exocyclic amino group changed to carbonyl oxygen and N1 protonated) was undetectable using this assay (a more than 2,000-fold increase in  $K_d$ ).

We determined the minimal region of this protein required for ATP-binding by deletion analysis. For this study we used DNA-tagged proteins generated by RNase treatment of mRNA-displayed proteins, leaving the protein attached to the short DNA linker<sup>5</sup>. For each construct, we measured the fraction of protein that bound to the column and eluted with ATP (Fig. 3c). This analysis defines a core domain of 45 amino acids sufficient for ATP-binding. Using these data, a second series of constructs were generated as MBP fusion proteins and thrombin-released deletion fragment proteins were used to measure  $K_d$  values (Fig. 3d). These solution binding experiments reveal a progressive loss of affinity as the deletion extends into the 70 N-terminal amino acids of the random region, from a  $K_d$  of 100 nM for the intact protein to 5  $\mu$ M for the 45-amino-acid core domain. The regions flanking the core domain might stabilize its structure, or might contribute additional distinct interactions with ATP.

These (family B) proteins contain two invariant CXXC sequences



**Figure 3** Sequences of selected ATP-binding proteins. **a**, Consensus sequences of the four selected protein families at round 8 (A, B, C, D), before mutagenesis. Flanking constant region residues are indicated by dashes. These protein sequences are unrelated to each other or to any known biological protein. **b**, The consensus sequence of the single remaining protein family at round 18 (18predom). Four predominating substitutions (>39/56 clones) and sixteen other selectively enriched (>4/56 clones) substitutions are also indicated (18select). Clone 18-19 had the lowest  $K_d$  for ATP (100 nM) of those assayed. **c**, Deletion analysis of clone 18-19 using DNA-tagged proteins<sup>5</sup>. Purified

constructs were incubated with ATP-agarose, washed and eluted with free ATP. The fraction present in the elution phase is shown, normalized with respect to the full-length DNA-tagged protein. Invariant residues are indicated by dashes. **d**, Deletion analysis of clone 18-19 using MBP fusion proteins. Successive deletion constructs of clone 18-19 were generated as MBP fusion proteins, then cleaved from the MBP with thrombin. Apparent  $K_d$ s were determined by displacement equilibrium filtration<sup>14</sup>. Invariant residues are indicated by dashes.

(56/56 clones sequenced) within the core domain. Also, protein 18-19 is functional in the presence of 5 mM dithiothreitol, indicating that disulphide bonds are probably not required for ATP binding. In addition, a deletion construct removing only one of these cysteines did not bind ATP (Fig. 3c). These observations suggested that family B proteins might require a coordinated metal ion to be functional. Elemental analysis by atomic absorption spectroscopy revealed the presence of one equivalent of bound zinc, and no other divalent metals. Incubation of the protein with EDTA results in a concentration-dependent loss of ATP-binding activity. Activity can be restored to protein that has been extensively dialysed in the presence of EDTA by the addition of  $\text{Zn}^{2+}$ , but not  $\text{Mg}^{2+}$ , ions.

We suggest that the family B proteins bind to ATP with a folded structure nucleated around, or stabilized by, a  $\text{Zn}^{2+}$  ion coordinated to the four invariant cysteines of the CXXC sequences. None of the other three protein families selected in this study contains this zinc-binding motif, and no known biological nucleotide-binding domain is a zinc-stabilized structure, although the glucocorticoid receptor contains a similar pair of CXXC sequences and binds to DNA in a zinc-dependent manner<sup>6</sup>. A NetBLAST<sup>7</sup> search of the NCBI protein sequence databases shows that the most closely related known protein sequence to the 45 amino acids of the minimal functional sequence of this protein has only a 33% identity. This is not a significant homology for a sequence this short, and did not include any of the four conserved cysteines. Zinc was not added to the selection buffer, but is present at about 10  $\mu\text{M}$  in mammalian blood<sup>8</sup>, from which the reticulocyte lysate used for mRNA translation is prepared. This result suggests that metal ion coordination may be one of the simplest ways of generating folded proteins while

minimizing the information required to specify a functional sequence.

The frequency with which ATP-binding proteins occur in sequence space can be estimated from the observed recovery of four such proteins from a non-redundant library of  $6 \times 10^{12}$  random sequences. On the basis of the average behaviour of the proteins isolated before mutagenesis (Fig. 2), only about 10% of the potentially functional sequences present in the first round would be expected to generate correctly folded active proteins and thus survive to be amplified. Detailed measurements of the efficiency of each step in the selection and amplification process confirmed that there were no significant material losses that would have affected the recovery of active proteins.

We therefore estimate that roughly 1 in  $10^{11}$  of all random-sequence proteins have ATP-binding activity comparable to the proteins isolated in this study. This frequency is similar to the recovery of ATP-binding RNAs from random-sequence RNA libraries (with similar  $K_d$  values)<sup>2,3</sup>. This suggests that the greater functional diversity of amino acids as compared with nucleotides does not make functional proteins more common than functional RNAs, at least with regard to binding small molecules. This may be due to the difficulty of forming a buried hydrophobic core that will fold in a consistent manner, in contrast to the ease with which nucleic acids can form compact structures stabilized by complementary base pairing. Although only a small proportion of each of the initially selected proteins folds into the active conformation, this proportion can be increased by directed evolution.

The combination of mRNA display with *in vitro* selection is a powerful approach to the exploration of protein sequence space. Our isolation of new functional proteins shows that it should be possible to obtain an unbiased view of the inherent diversity of all possible protein structures, and to determine whether biological proteins represent only a small subset of this diversity. Comparing the sequences of our newly evolved ATP-binding proteins with biological ATP-binding proteins has not revealed any significant similarity; structural data will also be required to reveal whether these proteins, especially the  $\text{Zn}^{2+}$  metalloprotein, are similar to those of any biological proteins.

In conclusion, we suggest that functional proteins are sufficiently common in protein sequence space (roughly 1 in  $10^{11}$ ) that they may be discovered by entirely stochastic means, such as presumably operated when proteins were first used by living organisms. However, this frequency is still low enough to emphasize the magnitude of the problem faced by those attempting *de novo* protein design. □

## Methods

### Library construction, *in vitro* selection and amplification

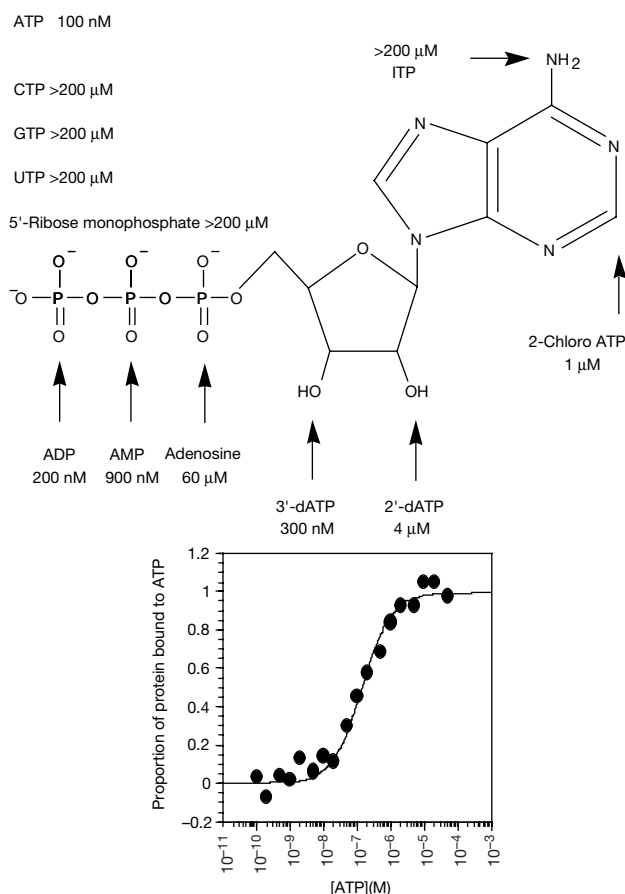
Library construction, and the preparation and purification of mRNA-displayed proteins have been described<sup>4,9,10</sup>. *In vitro* translation yielded  $7 \times 10^{13}$  independent random-sequence mRNA-displayed proteins; after purification and reverse transcription  $6 \times 10^{12}$  were recovered for input into the first selection step. Subsequent rounds were carried out at a 2–10-fold reduced scale. For rounds 1–9, we used a butyl-agarose pre-column (Sigma) and incubated the flowthrough with the ATP-affinity column. Rounds 14–16 included two ATP-agarose selection steps, and rounds 17 and 18 included three ATP-agarose selection steps. For reiterated selection steps the eluted material was purified away from ATP on a denaturing Ni-NTA column and reverse transcribed again before the subsequent selection step. For a more detailed description of the selection protocol and the full sequences of the selected proteins, see Supplementary Information.

### Mutagenic PCR amplification

Mutagenic PCR<sup>11,12</sup> was used in rounds 10–12, with an average mutagenic rate of 3.7% at the amino-acid level as determined by DNA sequencing. Serial transfer mutagenic PCR was carried out to an average mutagenic extent of 3.7% at the amino-acid level, with aliquots being combined to give a broad range of mutagenic extents.

### MBP fusion protein expression

Selected clones were ligated into pIADL14 and modified versions of pIADL14 (pTAG2K), and protein expression was carried out as described<sup>13</sup>. The proteins were further purified on a Ni-NTA column under denaturing conditions.



**Figure 4** Dissociation constants of protein-ATP (and ATP analogue) complexes as determined by displacement equilibrium filtration<sup>14</sup> for protein 18-19 as an MBP fusion protein.



## **K<sub>d</sub> determinations**

We made  $K_d$  measurements by equilibrium dialysis or spin-filtration<sup>14</sup>, with equivalent results, using purified MBP fusion proteins exchanged into selection binding buffer by gel filtration. We determined the  $K_d$  for ATP and ATP analogues by displacing bound <sup>32</sup>P-ATP from the protein by increasing concentrations of competitor. The stoichiometry of the protein-ATP interaction was determined by a procedure adapted from ref. 15.

## **Deletion analysis**

Deletion analyses were done with both DNA-tagged proteins<sup>5</sup> and MBP fusion proteins. Primers designed to anneal to specific internal parts of clone 18-19 were synthesized. For the DNA-tagged proteins, 5' primers encoded the TMV enhancer and T7 promoter sequences, with MD as the common N terminus, and 3' primers encoded the peptide MGMSG as the common C terminus. These primers were used to generate truncated mRNA-displayed proteins from clone 18-19 as described above. Subsequent purification on oligo(dT)cellulose, incubation with RNase A, and a second purification on oligo(dT)cellulose yielded purified DNA-tagged proteins<sup>5</sup>. The fraction of each DNA-tagged protein that was competent to bind ATP-agarose was determined under selection conditions as described above. MBP fusion proteins were generated in a similar manner by PCR of internal fragments of the protein 18-19 coding sequence, followed by cloning into the MBP fusion expression plasmid, expressing and purifying the MBP fusion deletion constructs, thrombin cleavage to release the deletion fragment from the MBP, and measuring solution  $K_d$  by equilibrium dialysis or spin-filtration.

## **Gel-filtration chromatography**

Protein 18-19 was expressed as an MBP fusion as described above. The time taken for the protein to traverse a Sepharose 200 FPLC column in selection binding buffer was measured on an AKTA FPLC by analysis of column fractions on silver-stained SDS-PAGE gels. The molecular weight of the protein was calculated by comparison with known protein standards.

Received 21 September 2000; accepted 16 January 2001.

- Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl Acad. Sci. USA* **94**, 12297–12302 (1997).
- Sassanfar, M. & Szostak, J. W. An RNA motif that binds ATP. *Nature* **364**, 550–553 (1993).
- Wilson, D. S. & Szostak, J. W. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–647 (1999).
- Cho, G., Keefe, A. D., Liu, R. L., Wilson, D. S. & Szostak, J. W. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. *J. Mol. Biol.* **297**, 309–319 (2000).
- Wilson, D. S., Keefe, A. D. & Szostak, J. W. *In vitro* selection of high affinity protein-binding peptides using mRNA display. *Proc. Natl Acad. Sci. USA* (in the press).
- Freedman, L. P. *et al.* The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature* **334**, 543–546 (1988).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Considine, D. M. (ed.) in *Van Nostrand's Scientific Encyclopedia* 7th edn, 3067 (Van Nostrand Reinhold, New York, 1989).
- Liu, R., Barrick, J., Szostak, J. W. & Roberts, R. W. Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. *Methods Enzymol.* **318**, 268–293 (2000).
- Keefe, A. D. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) Unit 24.5 (Wiley, New York, 2001).
- Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
- Wilson, D. S. & Keefe, A. D. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) Unit 8.3 (Wiley, New York, 2000).
- McCafferty, D. G., Lessard, I. A. D. & Walsh, C. T. Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. *Biochemistry* **36**, 10498–10505 (1997).
- Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).
- Wang, Y. & von Hippel, P. H. *Escherichia coli* transcription termination factor Rho. *J. Biol. Chem.* **268**, 13947–13955 (1993).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

## **Acknowledgements**

We thank members of the Szostak laboratory and especially D. Wilson, G. Cho, G. Short, J. Pollard, G. Zimmermann, R. Liu, J. Urbach and R. Larralde-Ridaura for their helpful advice. This work was supported in part by the NASA Astrobiology Institute and the NIH. J.W.S. is an investigator at the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to J.W.S. (e-mail: szostak@molbio.mgh.harvard.edu). The DNA sequences encoding the consensus protein sequences of families A, B, C, D, 18predom and clone 18-19 have been deposited in GenBank under accession codes AF306524 to AF306529, respectively.

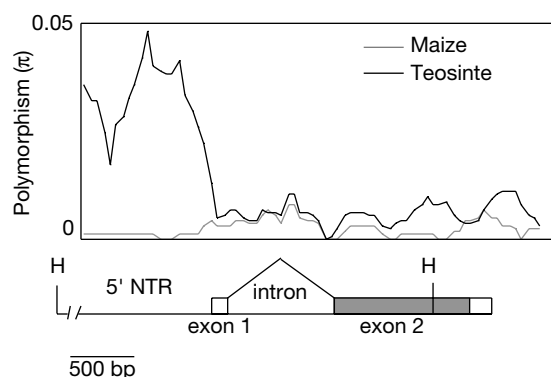
## **correction**

# The limits of selection during maize domestication

Rong-Lin Wang, Adrian Stec, Jody Hey, Lewis Lukens & John Doebley

*Nature* **398**, 236–239 (1999)

The primers used for PCR of the 3' portion of *tb1* amplified a duplicate locus (*tb1* homeologue) in two samples (11 and 16). Inclusion of these homeologous sequences caused  $\pi$  to rise sharply at the 3' end of the gene in Fig. 1. Using a new 3' primer (gagcatcatccagcagacgagaaa), *tb1* sequences were re-isolated (Genbank accessions AF340187–AF340209). The redrawn figure (below) still shows  $\pi$  rising on average, but not as sharply. Because of the known problems with PCR, all statistical tests in the paper were based on sequences isolated as  $\lambda$  clones, and the PCR-isolated sequences were used only in Fig. 1. Thus, other than Fig. 1, no statements or conclusions need amendment. An HKA test with the newly isolated 3' sequences shows no deviation from neutral expectations for maize ( $\chi^2 = 0.49$ ,  $P = 0.78$ ), and  $\pi(\times 1,000)$  for these sequences is 6.7 for teosinte and 5.8 for maize, confirming that selection is not apparent in the 3' region of *tb1*. □



## **erratum**

# Scabrous complexes with Notch to mediate boundary formation

Patricia A. Powell, Cedric Wesley, Susan Spencer & Ross L. Gagan

*Nature* **409**, 626–630 (2000).

The wrong symbol was placed next to Cedric Wesley's name in the author list. He is not presently at the National Science Foundation, but contributed equally to the work with Patricia A. Powell. □

# Going by the book

A heavy representation of catalogues in this week's selection.

## 2001 Biochemicals catalogue

**Roche** [www.roche.com/diagnostics](http://www.roche.com/diagnostics)

*Well ordered*

Laboratory reagents and other products for molecular biology, cell biology, immunochemistry, biochemistry and drug and disease research are included in this year's catalogue from Roche. A new section on instruments features the LightCycler for quantitative PCR and the Lumi-Imager F1 workstation for quantitative detection of chemiluminescence. Ordering has been simplified by the addition of an alphabetical product index as well as an index of products by catalogue number.

## 2001–2001 catalogue

**Vector Laboratories** [www.vectorlabs.com](http://www.vectorlabs.com)

*Banish the blues*

Vector's new catalogue lists nearly 500 products, including non-radioactive labelling and detection systems and Vectastain ABC kits. New products include Hematoxylin QS, a modified Mayer's haematoxylin designed to provide crisp, blue-violet nuclear staining, without a 'blueing' step, in less than 45 seconds. A new hydrophobic barrier pen, the ImmEdge, contains no ozone-depleting solvents and produces a heat-stable, water-repellent, lightly coloured barrier that keeps staining reagents localized on tissue section. Reagents for molecular biology include a new 5' EndTag nucleic acid labelling kit and biotinylated anti-maltose binding protein.

## Endogen catalogue

**Pierce** [www.piercenet.com](http://www.piercenet.com)

*Web-based quiz*

The Pierce Endogen Products Catalogue features hundreds of products for laboratories engaged in immunology, protein chemistry, proteomic and molecular biology research, including technical and ordering information for dialysis and gas chromatography products. The catalogue features a year-long trivia quiz. Answers to more than 100 science-related trivia questions scattered throughout the catalogue may be submitted to a weekly quiz on the company's website.

## SimplyBlue SafeStain

**Invitrogen** [www.invitrogen.com](http://www.invitrogen.com)

*Coomassie G-250 stain*

This ready-to-use stain does not require acid/alcohol fixatives or destains, and SimplyBlue can detect 1 µg of protein after just 5



Out of the blue: Invitrogen's simple protein stain.

minutes in the staining reagent. An optional water-based destaining step may be performed to achieve maximum sensitivity, to as low as 7 ng of protein.

## Laboratory Products

**Millipore Corporation**

[www.millipore.com](http://www.millipore.com)

*The 2001–2002 Laboratory Products catalogue*

Millipore offers thousands of membrane-based products for purifying and concentrating samples for molecular biology, protein chemistry applications, and assay development, as well as products for decontaminating fluids. Millipore manufactures products in facilities that comply with ISO and GMP requirements. The catalogue includes sample preparation products for genomics, proteomics, drug discovery and sterile filtration. New products featured in this catalogue include Montage Life Science Kits for nucleic acid research. The Montage Kits have been specifically developed to provide life scientists with complete and ready-to-use packages of components for select nucleic acid applications.

## Centrifuge accessories

**Nalge Nunc International**

[www.nalgenunc.com](http://www.nalgenunc.com)

*Tubes for all occasions*

The 2001 catalogue of Nalge Nunc Centrifuge Tubes and Bottles provides product infor-



mation on a wide selection of centrifuge tubes, bottles, closures and adapters, offering sizes and resin types to fit almost any assay requirement. Tubes are moulded from bioanalytical-grade resins. The range includes conical tubes, round-bottom tubes, Oak Ridge tubes and Ultratubes, all sized to fit the latest generation of lightweight carbon-fibre/composite rotors.

## BioWest

**UVP**

[www.uvp.com](http://www.uvp.com)

*Sensitive chemiluminescent detection*

This HRP-based, non-isotopic, horseradish peroxidase substrate kit is designed for enhanced chemiluminescent Western blot and ELISA detection, and is claimed to be 25 times more sensitive than other peroxidase-based kits. The duration and high signal intensity generated are easily identified and femtogram level detection visibly achieved. The kit is specifically formulated for use with CCD camera-based detection systems.

## Autokit CH50

**Wako**

[www.wakousa.com](http://www.wakousa.com)

*Liposome-based assay for total complement activity*

Autokit CH50 is a fully automated homogeneous liposome-based assay system for total complement activity in human serum. The complement cascade, consisting of approximately 20 serum proteins, plays an

## new on the market

important role in the body's immunological defence system. The assay is based on immune lysis of liposomes. A sample is mixed with the liposome and the substrate, the antibody in the reagent combined with dinitrophenyl on the liposomes and complements in the sample activated by the antigen-antibody complex. The activated complements break the membrane of the liposome. The enzyme glucose-6-phosphate dehydrogenase (G6PDH) contained in the liposome reacts with NAD and glucose-6-phosphate (G6P) in the reagent. During this enzyme reaction the NAD is reduced to NADH and, as a result, absorbance at 340 nm increases. Total precision studies give 3.2% CV and 5.7% CV at mean concentrations of 48.3 and 26.9 U per ml, respectively. The linearity of the assay is up to 65 U per ml, with a lower limit of detection of 10 U per ml.

### RNAqueous-96

**Ambion** [www.ambion.com](http://www.ambion.com)

*A new high-throughput RNA isolation kit*

Suitable for isolation of total RNA from cells or tissue in a high-throughput, 96-well plate format, this kit provides high yields of pure RNA free of contaminating RNases and genomic DNA and compatible with both real-time and end-point RT-PCR analysis. It can be used with many high-throughput vacuum manifolds and centrifuges.

### CISPro

**ChemSW** [www.chemsw.com](http://www.chemsw.com)

*Web-wise chemical inventory system*

This version of the company's Chemical Inventory system (CIS) provides access to real-time, company-wide chemical inventory information using Internet Explorer. The system eliminates the need to configure individual workstations and facilitates distribution and administration of the comprehensive chemical inventory system over widespread areas of the user company.

### Filtration plates

**Schleicher & Schuell** [www.s-und-s.de](http://www.s-und-s.de)

*Easy-to-use plates for fast filtration*

This new range of filter plates from S&S is marketed as a cost-effective solution for high filtration throughput for a range of applications including micro/macro array, cell harvesting, DNA separations and general sample clean. The plates incorporate a glass fibre support to enhance flow rate in filtration applications. They are available in a standard format, 96-well plate and can simultaneously filter 96 samples of up to 350 µl volume. The wells are individually sealed with an individual drainage spout which directs eluant into the collection plate wells, resulting in a significant reduction in



From S&S, rapid-filtration plates.

cross-talk, a contamination problem which can occur with other products.

### Amplifluor Apoptosis Gene Systems

**Intergen** [www.intergen.com](http://www.intergen.com)

*Good housekeeping*

The need for post-PCR manipulation is eliminated with these systems, which are based on incorporation of energy transfer-labelled hairpin primers into the amplification product. The primers are designed so that a fluorescent signal is generated when the primer is unfolded during its incorporation into an amplification product. This signal directly correlates with the accumulation of PCR product at each cycle. Unincorporated primers have a low fluorescence signal, allowing PCR and fluorescent signal detection to take place in a single reaction vessel, with no need to purify the PCR reaction prior to quantification. Either end-point or real-time fluorescence detection methods can be used. The systems detect mRNA expression of the apoptotic genes bcl-2, bax-a, fas and bclX (long). Housekeeping gene systems control for sample-to-sample variation in RNA quality and quantity and cDNA conversion efficiency by normalizing levels of amplified cDNA targets of interest to  $\beta$ -actin, GAPDH or 18S rRNA.

### Life Sciences

**Promega** [www.promega.com](http://www.promega.com)

*The 2001 catalogue*

The Promega Life Science catalogue 2001, now available, has been revised to include more comprehensive information, and to

integrate with the company's website for more detailed product and technical information. Literature, references and interactive troubleshooting zones can be found at [www.promega.com](http://www.promega.com) and the online catalogue and price list is updated daily with new and discontinued products.

### KillerTRAIL

**Alexis Biochemicals** [www.alexis-corp.com](http://www.alexis-corp.com)

*Killer application for manipulating cell death*

KillerTRAIL is a recombinant protein that induces apoptosis in a variety of TRAIL-sensitive cells (TRAIL is TNF-related apoptosis-inducing ligand). It binds to both human and mouse TRAIL receptors and does not require a cross-linker for biological activity. Anti-TRAIL MAb 2E5 neutralizes the activity of KillerTRAIL as well as other sources of soluble recombinant human TRAIL. The combination of KillerTRAIL and anti-TRAIL MAb 2E5 provides cancer researchers with a powerful tool for rapidly inducing and inhibiting cell death in TRAIL-related research

*These notes are compiled in the Nature office from information provided by the manufacturers.*

## ADVERTISEMENTS

### BACs • YACs • PACs cDNAs

We provide clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

**ResGen**<sup>™</sup>

Invitrogen Corporation

U.S. or Canada 800-533-4363  
FAX 256-536-9016

Check out our web site at  
[www.resgen.com](http://www.resgen.com)

**DIVERSE  
SMALL  
MOLECULES  
READY FOR  
SCREENING**



**10K-50K  
Compounds  
in Plates**

- Next Day Delivery\*
  - Competitively Priced
- \*(Some Restrictions Apply)

**Chembridge Corporation • 800.964.6143**  
858.451.7400 • 858.451.7401 fax • E-mail: [chem@chembridge.com](mailto:chem@chembridge.com)

**[www.chembridge.com](http://www.chembridge.com)**



**Bright future**

European synchrotron projects radiate strong career prospects

**p722**

**Sound structure**

Impending US boom in protein analysis calls for more beamlines

**p723**

**Money worries**

Are low salaries slowing the influx of scientists into synchrotrons?

**p723**

**Expertise wanted**

Structural genomics in Japan threatened by staff shortages

**p724**

# Increase in protein analysis pushes demand for synchrotron operators

From biology to physics, synchrotrons offer a bright future to users and specialists alike, says Helen Gavaghan.

**“**If I were a young scientist now, I would want to work with synchrotrons. The technical developments of these machines are extremely exciting and there is much virgin territory.” So says Giorgio Margaritondo, chair of the European Round Table for Synchrotron Radiation and Free Electron Laser, and a professor at the Swiss Federal Institute for Technology in Lausanne.

Synchrotrons are extremely versatile tools. They produce radiation across a broad range of wavelengths, allowing scientists to discern the atomic structure and behaviour of a variety of materials. As a user, Margaritondo might visit a synchrotron facility once or, if he were lucky, twice a year. Analysis of the data collected during such a visit could well keep a research team busy for the whole year.

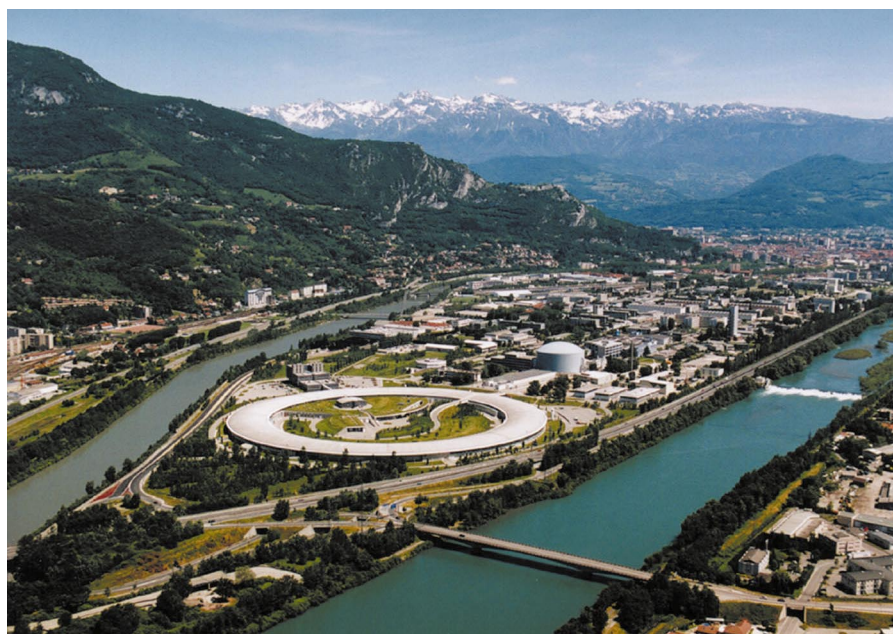
At the same time as reaping the immediate rewards of their experiments, says Margaritondo, scientists are identifying new ways of exploiting synchrotron beams. It is this territory that Margaritondo thinks thousands of scientists from many disciplines will find exciting.

During the past decade there has been an explosion of interest in synchrotron radiation and its uses. Accelerator physicists, beamline scientists and engineers have developed a deeper understanding of the phenomenon and how it can be manipulated.

But even as scientists are finding new uses for existing facilities, and fresh ways to squeeze more performance from them, new facilities are being built and even newer technologies are being proposed. So rapidly has the technology developed, says Margaritondo, that users from all disciplines and those operating the machines are scrambling to keep up with what the ever-advancing facilities have to offer.

**Inside a synchrotron**

At Daresbury, home to Britain's current national synchrotron facility, some 270 people work directly on the synchrotron with another 140 employed in computing, engineering and instrumentation design



**Light touch:** synchrotron facilities, such as the one based in Grenoble, are in demand from structural biologists because they can reveal the full structures of proteins.

and support. These numbers are typical for synchrotron facilities worldwide.

With the construction of the new synchrotron facilities in Britain and France — known as Diamond and Soleil, respectively — plus plans to add beamlines at Daresbury and keep the facility running until 2007, there will be exciting jobs to be had in the development and operation of these instruments. And in Britain, the government has announced several new projects in the north-east, including a proton synchrotron for medical applications and a fourth-generation free-electron laser light source.

The thousands of users queuing for beam time at synchrotrons would be lost without the accelerator physicists, beam scientists and engineers who provide expert support. And these staff members often undertake research projects themselves.

Synchrotrons are a uniquely multidisciplinary environment, says Susan Smith, a senior accelerator physicist at Daresbury.

Smith has a degree in physics and maths from the University of Glasgow and did graduate work on free-electron lasers before joining Daresbury in 1985. In the past few years, her time has been split between providing support for Daresbury's synchrotron radiation source, doing generic accelerator studies and working on new light sources.

Rob Lewis has designed and worked with detectors since 1979 for the radiation source at Daresbury. He is now heading a science team to develop medical imaging as part of the proposed proton synchrotron to be based at the same site. Lewis has a degree in physics, a PhD in X-ray astronomy and post-doc experience in the same field. He says that he has seen quite a lot of people with his academic background, and an experimental rather than theoretical inclination, move as he has done towards medical physics.

Understanding synchrotron radiation or particle accelerators in depth is not a prerequisite for the work, says Lewis. When he

ESRF

## What is a synchrotron?

Ironically, synchrotron radiation, although now greatly in demand, was a bane in the life of high-energy physicists working with particle accelerators. This is because the radiation represents a loss of energy.

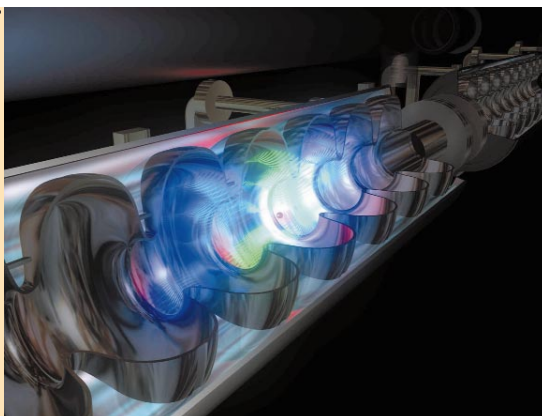
In particle accelerators, charged particles travel millions of times round a doughnut-shaped ring containing a vacuum. Strategically placed magnets confine the particles to their circular path. The strength of the magnetic field must change as the particles accelerate, in order to keep them on course. Because this change must be synchronized with the energy gain associated with increasing the particles' speed, the radiation emitted acquired the name synchrotron radiation.

Particle accelerators are designed to compensate for the energy loss, but once scientists saw that synchrotron radiation might be useful, machines were built specifically to produce it — the 'second-generation' facilities. In these systems, the particle beams are confined in a 'storage ring' and siphoned off in a controlled manner along a beamline.

Instruments in the beamlines, such as monochromators, determine the frequencies available for a particular class of experiments — perhaps probing for fault lines or impurities in a material, or X-ray diffraction to determine a protein structure.

As accelerator physicists learned more about synchrotron radiation, they designed magnets engineered with precise proportions to increase the energy and brightness in a beam and so expand its potential range of applications. Instead of squeezing these 'insertion devices' into synchrotrons as an afterthought, 'third-generation' synchrotrons are specially designed with the devices in mind. One such facility, the Swiss Light Source, is scheduled to come online this August in Villigen, and Diamond and Soleil are planned for later this decade in Britain and France.

H. G.



ease this situation, users and providers of synchrotron radiation need to find ways to exploit both the second- and third-generation facilities more effectively.

In particular, the beamlines at these synchrotrons need better detectors, which offers fresh opportunities for beamline scientists and engineers. "In many cases it's not brightness that limits experiments, but how fast the detectors can record," says Lewis. Margaritondo agrees that there is a need for a major international push to develop better detectors.

But even as the third-generation facilities spring into operation around the world, the European Round Table for Synchrotron Radiation and Free Electron Laser is pondering the next step. There are two options: design a synchrotron at the limit of what physics allows or move to an alternative technology — free-electron lasers — that could surpass even the most advanced synchrotron.

One possibility could be for the European Synchrotron Research Facility based in Grenoble to develop the 'ultimate' synchrotron. Meanwhile, TESLA, a German proposal for a free-electron laser that can produce high-energy X-rays, would also be developed with international collaboration. "I sense a consensus developing that we should develop both," says Margaritondo.

So, whether as a user developing new techniques to exploit beamlines, as an accelerator physicist, beamline scientist or engineer, the job prospects at synchrotron facilities are bright for some time to come. And it is hard to imagine that in a decade or so any branch of science will not have some need of synchrotrons.

**Helen Gavaghan is a freelance science and technology journalist based in Hebden Bridge, West Yorkshire.**

Synchrotron radiation facilities

♦ [http://www.src.wisc.edu/links/sr\\_sources.html](http://www.src.wisc.edu/links/sr_sources.html)

European Synchrotron Research Facility

♦ <http://www.esrf.fr>

European Round Table for Synchrotron Radiation and Free Electron Laser

♦ <http://www.elettra.trieste.it/sites/roundtable>

joined Daresbury he knew nothing about the subject. Like Smith and others who join synchrotron facilities around the world, he learned to apply his knowledge of physics to the subject once he had got the job.

Typically, when scientists such as Lewis join national synchrotron facilities they become part of a team that deals with a class of experiments, such as protein crystallography. They are likely to be physicists, chemists or engineers by background. "Anything from 80% upwards of your time is supporting the users. You need to understand what they want to achieve and to liaise with them about what the equipment can do," Lewis says. "There is a real dearth of people who can fill this middle ground."

### Beam-time shortage

Although the next generation of machines (see 'What is a synchrotron?', above) will mean there are more beamlines available, competition among users for beam time is, and will continue to be, acute, says Guy Le Lay, a professor at the University of Provence and member of the European Round Table for Synchrotrons and Free Electron Laser. "Obviously some competition — perhaps 150% over-subscription — is healthy," says Margaritondo, who chairs the round table, "but beamlines currently are between 200% and 300% oversubscribed and the situation for protein crystallography is even worse."

Given that opportunities are now opening up for biologists to learn about the proteins that the sequencing of the human genome tells them the body is producing, the pressure is not surprising and it will not ease up. But it is not only the biologists that are clamouring for time on synchrotron beamlines. The X-rays produced in synchrotron facilities are used by chemists, materials scientists and physicists alike.

Jörgen Albertsson, professor of inorganic chemistry at Chalmers University of Technology in Sweden, says that it can be hard to compete successfully against the protein crystallographers for beam time. Although the construction of Diamond and Soleil will

## Automated for the people

Without new synchrotrons, US structural genomics may depend on automating existing facilities to meet its goals. But without an influx of interested physicists and engineers — and the funding to support them — speeding up the process of discerning a protein's three-dimensional structure may prove difficult.

The automated gene sequencer revolutionized genomics. But no single machine, by itself, will greatly speed up structural biology. Working out a protein's structure is a more complex process, with many more stages, says Peter Kuhn, assistant professor at the Stanford Synchrotron Radiation Laboratory.

The protein must be purified, crystallized, mounted in the beamline, probed by the light, then reconstructed by computer. Potential bottlenecks occur at every stage — and each stage may need different combinations of expertise to speed it up. For example, automating sample mounting in the beamline requires engineering skills, whereas finding and centring the sample within the beam needs computational skills.

Different US groups are working on separate stages of this problem. For example, Stanford is in the nascent stages of automating sample mounting. Once individual steps have been automated, systems engineers will need to evaluate how the entire process fits together, Kuhn says.

P. S.



# US structural genomics effort needs physicists for success

**A**n impending boom in structural genomics is placing increased demands on US synchrotrons — promising jobs for physicists willing to work with biologists in deciphering the three-dimensional shapes of thousands of proteins.

In the United States, the field is being catalysed by a five-year, \$150-million initiative by the National Institute of General Medical Sciences (NIGMS). This pilot programme is funding seven consortia to find the fastest and best ways to solve thousands of structures (see *Nature* **407**, 549; 2000). If successful, NIGMS officials predict they could decipher between 10,000 and 30,000 structures in 10 years — representing most families of protein folds that exist in nature.

But for this ambitious experiment to succeed, structural biologists will have to rely on, and work with, physicists. This is especially true in the case of synchrotron light sources. These facilities produce intense, highly focused light that can be used to probe the atomic structure of proteins and other materials. Each synchrotron siphons off light produced by the high-speed particles into several 'beamlines', each of which can be equipped with different detectors to probe the properties of various materials. Synchrotrons are considered to be one of the best tools for deciphering protein structures.

At the moment, physicists are in demand for developing new facilities, staffing existing ones, and finding ways to automate many aspects of the protein structure discovery process. But there are only five national synchrotron light sources in the United States — less than half the number in Europe. And many US structural biologists and synchrotron staffers say there is not enough US synchrotron capacity to support existing work, much less the high volumes needed to make the kind of dent in the proteome the structural genomics initiative seeks. "There aren't enough beamlines for it to grow fast enough," says Bob Sweet, a structural biologist at Brookhaven National Laboratory's National Synchrotron Light Source. "There's a ceiling. We're going to hit that ceiling in three to five years if we don't get more beamlines."

Although there are no solid proposals to build new synchrotron facilities in anticipation of a structural-biology boom, there are several plans at various stages of approval to add more beamlines devoted solely to structural biology. Those initiatives call for adding between 12 and 15 biology beamlines in the next five years to the 40 or so already dedicated to the discipline. If those beam-

lines get approval, several physicists will be needed to develop, and then staff, each of them.

Sweet suspects that adding beamlines to existing facilities may be a short-term fix at best. New facilities may be needed, but the decision will have to be made soon to proceed, as construction of a new synchrotron light source takes up to six years. "We must decide now what we need in six to ten years," he says. "Otherwise we'll be in the dark."

Kevin D'Amico, who has about 20 years of synchrotron experience, now earns a living as a consultant on how to build beamlines devoted to biology. He is currently working with Structural GenomiX, a San Diego biotech company that is building its own beamline at Argonne National Laboratory's Advanced Photon Source (APS).

D'Amico sees the niche for beamline developers as fairly narrow, with demands in the tens, rather than the hundreds. Qualifications include spending several years working at a synchrotron facility — the same qualifications needed for someone to run either a beamline or an entire facility.

## Staff shortage

Gerd Rosenbaum, director for the South-eastern Regional Collaborative Access Team at Argonne, agrees. "The demand will be greater for people to staff and operate beamlines," says Rosenbaum, who is overseeing development of a new beamline at the APS. "We have a serious shortage of trained staff at all beamlines — not just the biological ones."

Rosenbaum notes that the profession of X-ray crystallographer as a speciality in structural biology is dying out, as the tools for using X-rays to study protein structures are becoming easier for a wider community to use. But to make that tool effective, facilities need to attract and retain highly skilled, multidisciplinary teams of people. Teams that support crystallography are made up of structural biologists, computational experts and physicists.

All of these people must, of course, be well-versed in their own discipline, as well as able to communicate with their colleagues in other fields. For example, a physicist employed at a biology beamline must master the instrumentation associated with the experiments as well as understand the intricacies of scanning different kinds of proteins.

Having well-trained staff in place at each beamline will make the operation more efficient when outside users come in to do experiments, or when more users take the 'Fed-X-ray crystallography' approach. In



## Wages may make it hard to attract scientists to synchrotrons

Low salaries and high housing costs — this combination is making recruiting synchrotron staff difficult at some facilities in the United States and Britain.

Finding physicists and engineers to work at the Stanford Synchrotron Radiation Laboratory (SSRL) is an uphill battle because the university cannot offer wages that match those of companies in Silicon Valley, says Peter Kuhn, assistant professor at the SSRL. "You try to attract the best scientists and, at the same time, you're competing with private industry." That competition turns into a losing battle when potential staffers face the inflated value of real estate in Palo Alto.

The competition is further imbalanced when synchrotrons lack facilities. At the SSRL, beamlines are being added faster than offices can be constructed, so many staff are housed in trailers near the ring. Eventually, the SSRL will have the offices it needs, as the university builds the infrastructure needed to support its expanding interests in biology.

P. S.



▶ that approach, users mail in their samples, beamline staff generate the data, and the users analyse it remotely in real time, communicating with the synchrotron staff about whether or not more reads are necessary or if any adjustments need to be made.

Rosenbaum says institutional or cultural obstacles now make it difficult to attract and retain synchrotron staffers. Beamline operators, he says, are "consumed" by their users, and seldom work on their own projects. As a result, many find it hard to get first authorship on papers, making it difficult for them to earn the credentials necessary for a tenure-track position later in their careers. Without a stable staff, beamlines do not run as efficiently as they should, he says. Rosenbaum says the problem applies to synchrotrons throughout the world and extends beyond structural biology.

Sol Gruner, director of the Cornell High Energy Synchrotron Source (CHESS), says that his facility also has several openings.

Gruner notes that staff support at the facilities is more important than having the latest technology.

Although Gruner recruits computational specialists, structural biologists and physicists, he says the physicists—the people responsible for generating the light and seeing that it hits the sample—are the hardest to find.

That difficulty could reflect a general shortage of physicists in the United States. There are fewer physicists looking for work than biologists or chemists (see *Nature* 409, 445; 2001). The shortage may be especially pronounced at synchrotrons because qualified people must possess multidisciplinary skills. "You're talking about a small group of people who are very much in demand for their versatile talents," says Gruner. ■

Paul Smaglik is *Naturejobs* editor.

♦ <http://biosync.sdsc.edu>

♦ <http://www.nigms.nih.gov/funding/psi.html>

♦ <http://www.x12c.nsls.bnl.gov/StrGen.htm>

## Staffing shortage threatens Japan's structural genomics

**S**ynchrotron radiation facilities in Japan are suffering from a chronic staff shortage. Despite considerable funding increases in some application areas—such as structural genomics and nanotechnology—demand for beamline specialists, experimentalists, technicians and specialized software engineers is outstripping supply.

The bold move a few years ago to put structural genomics at the top of the agenda of the newly established RIKEN Genomic Sciences Center in Yokohama caused a stir within Japan's genetics research community. Although critics of the new centre pointed out that more emphasis should be put on genome sequencing, the centre's designated director, Akiyoshi Wada, argued that focusing efforts on structural genomics would constitute a more viable strategy for a "late-comer" in the genomics field.

With SPring-8, the world's largest synchrotron radiation facility, opening its doors for research in October 1997 in Harima, the timing for a structural genomics initiative was well chosen. A third-generation synchrotron radiation facility in the hard X-ray range, SPring-8 was planned and constructed jointly by RIKEN and the Japan Atomic Energy Research Institute (JAERI).

RIKEN presently occupies two of the four beamlines dedicated to structural biology at SPring-8. Using 'multiwavelength anomalous diffraction' and a device called a trichromator, Japanese scientists have successfully tackled the structural basis of glutamate receptors in the central nervous system, and the structure of various proteins includ-

ing bovine rhodopsin.

Funding for structural genomics has been exploding in Japan over the past two years. But, according to insiders, virtually all synchrotron radiation facilities remain inadequately staffed and user support is limited. According to Masashi Miyano, who directs the structural biology group at RIKEN's Harima research centre, the number of beamline staff at SPring-8 is about one-third that of the staff at comparable facilities outside Japan.

Automation may eventually provide part of the answer—especially for ambitious projects in structural genomics. One initiative, headed by Seiki Kuramitsu, a professor in the graduate school of science at Osaka University, aims to elucidate the structure and function of all proteins in the extreme thermophile *Thermus thermophilus*. For this project to

succeed, automating the whole process, from protein production to crystallization and data acquisition, will become crucial because of the large number of proteins.

Beyond increasing automation and hiring more people, Japanese synchrotrons could be improved by focusing more on service, several users say. There is widespread agreement that both existing and future synchrotron research facilities in Japan must spend more time reflecting the needs of their users and prioritizing beam-time allocation. Some facilities within Japan are well known for their lack of user support. And users from the biology community are increasingly demanding standardization of the equipment and ease of use.

Physicists such as Tsuneaki Miyahara, a professor in the department of physics at the Tokyo Metropolitan University, disagree, arguing that it is important to teach scientists how to improve instrumentation, rather than just how to use it. "Younger scientists generally do not want to go into the details of instrumentation, which might crucially impact on the future development of synchrotron radiation research in Japan," says Miyahara.

But it is not just hardware specialists that are needed. Structural biologists spend most of their time on model development and verification, says Miyano. The software tools that are needed for this purpose are coming mostly from the United States. "Without a long-term strategy to develop adequate software capacities, the situation seems critical," says Miyano. ■

Robert Triendl is a freelance writer based in Tokyo.

### Web links

SPring-8 ♦ <http://www.spring8.or.jp>

RIKEN Structural Genomics Initiative

♦ <http://www.rsgi.riken.go.jp>

KSR ♦ <http://www.wal.kuicr.kyoto-u.ac.jp>

KEK Photon Factory ♦ <http://www.kek.jp>

UVSOR ♦ <http://vsor-ntserver.ims.ac.jp>

HISOR ♦ <http://www.hsrc.hiroshima-u.ac.jp>

Ritsumeikan University SR Center

♦ <http://www.ritsumei.ac.jp/se/d11/index-e.html>

New Subaru

♦ <http://www.lasti.himeji-tech.ac.jp/NS/Index.html>

## NMR versus synchrotron radiation

The completion of the world's largest nuclear magnetic resonance (NMR) facility at the campus of the RIKEN Genomic Sciences Center in Yokohama is an important step forward for Japan's structural genomics programme. NMR spectroscopy is a potential rival to synchrotron radiation in protein-structure analysis and, with RIKEN's new facility, Japanese scientists hope to play a key role in developing new instrumentation tools for structural genomics. Initially, the new facility will have 16 machines, and this number is expected to increase to 200 at its completion in 2002. Among others, the facility will host the world's most powerful NMR device.

Although some scientists are sceptical about whether NMR spectroscopy will ever replace crystallography as a research tool in structural biology, others argue that much could be done to improve NMR techniques, and that experimental performance could surge with the availability of new, more powerful instruments.

Yoshiyuki Yokoyama, who is responsible for RIKEN's structural genomics effort, believes that X-ray crystallography and NMR spectroscopy have their perspective advantages and that combining both techniques is likely to be the most beneficial strategy. "X-ray crystallography and NMR spectroscopy are complementary technologies," he says.

R. T.